

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

**TAXONOMY AND VISUALIZATION OF DIGITAL ARCHITECTURE
KNOWLEDGE: PROPOSAL FOR A SCIENTIFIC ONLINE ENCYCLOPEDIA**



Ph.D. THESIS

Esratur KARACİF

Department of Informatics

Architectural Design Computing Programme

JULY 2025

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

**TAXONOMY AND VISUALIZATION OF DIGITAL ARCHITECTURE
KNOWLEDGE: PROPOSAL FOR A SCIENTIFIC ONLINE ENCYCLOPEDIA**



Ph.D. THESIS

**Esratur KARACİF
(523192008)**

Department of Informatics

Architectural Design Computing Programme

Thesis Advisor: Assoc. Prof. Dr. Ethem GÜNER

JULY 2025

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

**DİJİTAL MİMARLIK BİLGİSİNİN TAKSONOMİSİ VE
GÖRSELLEŞTİRİLMESİ:
BİLİMSEL BİR ÇEVİRİM İÇİ ANSİKLOPEDİ ÖNERİSİ**

DOKTORA TEZİ

**Esratur KARACİF
(523192008)**

Bilişim Anabilim Dalı

Mimari Tasarımda Bilişim Programı

Tez Danışmanı: Doç. Dr. Ethem GÜRER

TEMMUZ 2025

Esranur KARACİF, a Ph.D. student of ITU Graduate School student ID 523192008 successfully defended the dissertation entitled “TAXONOMY AND VISUALIZATION OF DIGITAL ARCHITECTURE KNOWLEDGE: PROPOSAL FOR A SCIENTIFIC ONLINE ENCYCLOPEDIA”, which she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor : **Assoc. Prof. Dr. Ethem GÜRER**
Istanbul Technical University

Jury Members : **Prof. Dr. Leman Figen GÜL**
Istanbul Technical University

.....
Prof. Dr. Neşe ÇAKICI ALP
Kocaeli University

.....
Assoc. Prof. Dr. Orkan Zeynel GÜZELCİ.....
Istanbul Technical University

.....
Assoc. Prof. Dr. Serdar AYDIN
T.C. Mardin Artuklu University

Date of Submission : 12 June 2025

Date of Defense : 4 July 2025



FOREWORD

The intellectual impetus for this dissertation arose from a direct, personal challenge encountered during my doctoral qualifying examinations. As I navigated the vast and complex landscape of digital architecture, I found that while information was abundant, coherent understanding was elusive. The struggle to connect disparate concepts and discern the underlying structure of a rapidly evolving field sparked a deeper inquiry: could we develop more effective models, not just to find information, but to comprehend its intricate relational fabric? This question guided my research toward envisioning a hierarchical taxonomy as a cognitive map, a tool to reveal the latent trends and patterns often obscured in linear texts. My aspiration was to create more than a repository of data; it was to build an environment for exploratory learning that could guide learners through the intellectual landscape I had found challenging. It is my sincere hope that the framework proposed in this thesis proves useful to the scholars and students who follow, making their own journeys of discovery clearer and more insightful.

This work would not have been possible without the guidance, support, and wisdom of many individuals. First and foremost, I extend my gratitude to my thesis advisor, Assoc. Prof. Dr. Ethem GÜRER, for his insightful guidance throughout this research.

I would also like to extend my heartfelt thanks to the esteemed members of my dissertation committee: Prof. Dr. Leman Figen GÜL, Prof. Dr. Neşe ÇAKICI ALP, Assoc. Prof. Dr. Orkan Zeynel GÜZELCI, and Assoc. Prof. Dr. Serdar AYDIN for their dedication, academic insight and constructive feedback, which significantly enriched this work.

My thanks go as well to all the participants who contributed to the empirical component of this research by responding to the surveys and sharing their perspectives.

I am especially grateful to my family for their belief in me and for providing emotional strength during the most demanding periods of this journey. To my close friends, thank you for your constant encouragement, patience, and understanding.

Finally, my deepest and most heartfelt thanks go to my beloved spouse. Your patience, empathy, and steadfast support have been my greatest source of strength throughout this long and challenging path. I could not have completed this journey without your constant presence and unwavering support.

July 2025

Esranur KARACİF
(Architect)



TABLE OF CONTENTS

	<u>Page</u>
FOREWORD	vii
TABLE OF CONTENTS	ix
ABBREVIATIONS	xi
SYMBOLS	xiii
LIST OF TABLES	xv
LIST OF FIGURES	xvii
SUMMARY	xix
ÖZET	xxi
1. INTRODUCTION	1
1.1 Statement of the Problem	1
1.2 Purpose, Scope and Method of the Thesis	4
2. EXPLORING THE IMPACT OF KNOWLEDGE STRUCTURING	9
2.1 Effects of Knowledge Categorization on Learning	9
2.2 Knowledge Types	12
2.2.1 Conceptual knowledge	16
2.2.2 Procedural knowledge	17
2.2.3 Visual knowledge	18
2.2.4 Configurational knowledge	19
2.3 Representation and Mapping of Knowledge	20
2.3.1 Visual representation of knowledge	20
2.3.1.1 Creative mapping in architecture theory	21
2.3.1.2 Visual-spatial mapping of knowledge	24
2.3.2 Textual representation of knowledge	25
2.3.2.1 Printed encyclopedias and dictionaries	26
2.3.2.2 Online encyclopedias and dictionaries	26
2.4 Mapping for Learning and Research in Online Hypermedia: Digital Humanities	30
2.5 Exploratory Learning via Interactive Data Visualization	32
3. ARTIFICIAL INTELLIGENCE-ASSISTED TEXT MINING	35
3.1 Text Preprocessing	37
3.2 Hierarchical Text Categorization and Clustering of Preprocessed Text Data	38
3.3 Text Mining Applied to Scientific Studies	43
4. DATA VISUALIZATION AS A LEARNING TOOL: MAPPING AND CATEGORIZATION OF DIGITAL ARCHITECTURE KNOWLEDGE	49
4.1 Knowledge Classification in Digital Architecture	49
4.2 Data Visualization as a Learning Tool	56
4.3 Phase 1: Testing Text Classification and Hierarchical Visualization with a Sampled Dataset	61
4.3.1 Retrieval of the raw data	62
4.3.2 Preprocessing of the data set	63

4.3.3 Categorization of the preprocessed data set	65
4.3.4 Visualization of the categorized data: Comparative analysis of hierarchical diagrams	66
4.3.4.1 First group: Zoomable, radial, solid and hierarchical representation.	67
4.3.4.2 Second group: Radial, linear, network-based and hierarchical representation	69
4.3.5 Web interface for the classification and visualization of digital architecture knowledge	71
4.4 Phase 2: Testing Text Classification and Hierarchical Visualization with a Full Dataset	74
4.4.1 Creation of RIS to JSON converter and categorized hierarchical keyword classification plug-ins.....	74
4.4.2 Zoomable sunburst diagram and prototype of the data visualization.....	80
4.4.3 Other elements of the web page	84
5. EVALUATION AND RESULTS ON TEACHING DESIGN OF DIGITAL ARCHITECTURE KNOWLEDGE: USER FEEDBACK ANALYSIS.....	91
6. DISCUSSIONS AND LIMITATIONS	105
6.1 Evaluating the Research Objectives	105
6.1.1 The potential of hierarchical structures for exploratory learning.....	105
6.1.2 The role of systematic classification and relational visualization.....	106
6.2 System Limitations and Data Challenges.....	106
6.3 Future Directions and Research Potential	108
7. CONCLUSIONS.....	111
REFERENCES.....	117
APPENDICES	125

ABBREVIATIONS

AI	: Artificial Intelligence
BIM	: Building Information Modeling
CAD	: Computer Aided Design
CAM	: Computer Aided Manufacturing
CD	: Computational Design
CUMINCAD	: Cumulative Index Of Publications About Computer Aided Architectural Design
ITU	: Istanbul Technical University
MBL	: Mimari Tasarımda Bilişim (eng. Architectural Design Computing)
ML	: Machine Learning
SUS	: System Usability Scale
TF-IDF	: Term Frequency-Inverse Document Frequency



SYMBOLS

$TF_{i,j}$	: TF value of term “i” in document “j”
n_{ij}	: Frequency of term “i” in document “j”
$\sum_k n_{k,j}$	: Total frequency of all terms in document “j”
IDF_i	: Inverse Document Frequency (IDF) value of term “i”
$\log$	: Logarithm function
n_d	: Total Number of Documents
$d: t_i \in d$	: Number of documents containing term



LIST OF TABLES

	<u>Page</u>
Table 2.1 : Classification of knowledge types according to different authors.....	13
Table 4.1 : Descriptions, advantages and disadvantages of first group of diagrams.	68
Table 4.2 : Descriptions, advantages and disadvantages of second group of diagrams.	70





LIST OF FIGURES

	<u>Page</u>
Figure 1.1 : The flowchart of the methodology.	6
Figure 2.1 : The hierarchical propositional model of Quillian (McClelland and Rogers, 2003).	11
Figure 2.2 : Connectionist model, parallel distributed processing network (McClelland and Rogers, 2003).	12
Figure 2.3 : Map of evolutionary architecture by Charles Jencks (Jencks, 1971). ...	22
Figure 2.4 : Richard Giblett, 21. mycelium rhizome (Giblett, 2009; Menken, 2019).	23
Figure 2.5 : Representation of the relations of the word "knowledge" with a network model on the Visuwords platform (Url-1).	27
Figure 2.6 : Representation of the relations of the word "knowledge" with a network model on the Visual Thesaurus platform (Url-2).	27
Figure 2.7 : Representation of the linguistic relationships with a network model on the Lexipedia platform (Url-3).	28
Figure 2.8 : Mapping the “design” discipline to build a learning architecture by Ksenia Cheinman (Url-4).	29
Figure 2.9 : Data visualization on CiteSpace platform (Url-5).	29
Figure 2.10 : Mapping on AR mapping created by the author in graph commons platform (Url-6).	30
Figure 2.11 : A network representation obtained via VOSViewer software (Url-7).	31
Figure 3.1 : The three steps of text indexing (Jo, 2019).	37
Figure 3.2 : A decision tree for identifying application areas in accordance with text mining resources (Miner et al., 2012).	39
Figure 3.3 : A Venn diagram illustrating the overlap between text mining and six associated disciplines (Miner et al, 2012).	44
Figure 3.4 : The underlying concept of Latent Dirichlet Allocation (LDA) (Blei, 2012).	45
Figure 4.1 : Data visualization examples of D3.js (Url-8).	60
Figure 4.2 : The flowchart of the preprocessing steps and data visualization (Po et. al., 2020).	60
Figure 4.3 : The general flowchart of the methodology of the first phase with sampled data.	61
Figure 4.4 : The detailed flowchart of the methodology of the first phase with sampled data.	62
Figure 4.5 : Extracting data from CUMINCAD database (Url-9).	63
Figure 4.6 : “.ris” file format retrieved from the CUMINCAD database.	63
Figure 4.7 : Raw data extraction from CUMINCAD and transformation into structured data after reduction and preprocessing.	64
Figure 4.8 : Diagram illustrating categorization of structured data using TF-IDF and k-Means clustering, followed by export to Excel.	65

Figure 4.9 : Visualization of categorized data and model building using two sets of data visualization tools.	67
Figure 4.10 : Zoomable, circular, solid, and hierarchical diagrammatic representations (From left to right: zoomable sunburst, zoomable circle packing, sequences sunburst).	68
Figure 4.11 : Radial, linear, and hierarchical diagrammatic representations (From left to right: hierarchical edge bundling, radial tree, tree of life).	70
Figure 4.12 : Zoomable sunburst diagram in the context of the web page.	72
Figure 4.13 : The elements of zoomable sunburst diagram in the web page.	72
Figure 4.14 : Conceptual (a), procedural (b), visual (c), and configurational knowledge (d) layer of the article (on the right side of the webpage) and the PDF of the article in an iframe on the webpage (left).	73
Figure 4.15 : Subcategories of Robotic Fabrication (left) and article with four knowledge layers (right).	73
Figure 4.16 : Conceptual, procedural, visual, and configurational knowledge layer of the article “Collaborative Robotic Masonry and Early Stage Fatigue Prediction” inside the zoomable sunburst diagram.	74
Figure 4.17 : RIS to JSON converter and hierarchical clustering of JSON data on the homepage.	75
Figure 4.18 : The output of the RIS to JSON converter on the homepage when a .ris file is uploaded.	77
Figure 4.19 : The output of the categorized hierarchical clustering of JSON data on the homepage.	79
Figure 4.20 : Zoomable Sunburst diagram created with D3.js for data visualization.	81
Figure 4.21 : Zoomable Sunburst Prototype diagram created with D3.js for data visualization.	81
Figure 4.22 : The “Conceptual Knowledge” represented by a hierarchical word cloud.	82
Figure 4.23 : The “Procedural Knowledge” represented by a flowchart.	83
Figure 4.24 : The “Visual Knowledge” represented by Microsoft Designer.	83
Figure 4.25 : The “Configurational Knowledge” created via GoJS.	84
Figure 4.26 : The navigation bar of the web page.	84
Figure 4.27 : The guideline on the usage of the web page and the questionnaire (in Turkish and English).	85
Figure 4.28 : About page detailing the website and the published articles.	86
Figure 4.29 : Zoomable Circle Packing diagram created with D3.js for data visualization.	87
Figure 4.30 : Sequences Sunburst diagram created with D3.js for data visualization.	88
Figure 4.31 : Hierarchical Edge Bundling diagram created with D3.js for data visualization.	88
Figure 4.32 : Radial Tree diagram created with D3.js for data visualization.	89
Figure 4.33 : Tree of Life diagram created with D3.js for data visualization.	90
Figure 5.1 : Graphical representation of SUS results.	99
Figure 5. 2 : Heatmap results from group 1 on the homepage (RealEye, 2024).	101
Figure 5. 3 : Heatmap results from group 2 on the homepage (RealEye, 2025).	102
Figure 5. 4 : Heatmap results from group 3 on the homepage (RealEye, 2025).	103

TAXONOMY AND VISUALIZATION OF DIGITAL ARCHITECTURE KNOWLEDGE: PROPOSAL FOR A SCIENTIFIC ONLINE ENCYCLOPEDIA

SUMMARY

This study introduces a comprehensive framework for the systematic taxonomy and visual mapping of digital architecture knowledge through the proposal of an AI-assisted scientific online encyclopedia. By integrating methodologies from digital humanities, architectural computing, computational design and knowledge representation, the research aims to reorganize domain-specific knowledge by first constructing a hierarchical structure and subsequently classifying this knowledge into four primary categories: conceptual, procedural, visual and configurational. This classification provides a multi-dimensional perspective on digital architecture discourse, establishing a pedagogical model designed to support exploratory learning. The study is organized into two main phases. The first phase involves experimentation with a sampled dataset in order to test classification methods and evaluate different visualization approaches. It begins with the extraction of metadata from the CUMINCAD database, which serves as a comprehensive repository of digital architecture publications. The dataset is processed through a series of natural language processing (NLP) procedures, including text preprocessing, term weighting via TF-IDF vectorization, and hierarchical clustering. These methods are applied to uncover latent thematic structures and conceptual relationships within the data. In the second phase, generative AI models are utilized to further classify and interpret the data. This phase moves beyond static algorithmic categorization and enables more nuanced, context-sensitive classifications. In this phase, the dataset is significantly expanded to include the full corpus of digital architecture publications. The increased scale and complexity of the data necessitate more advanced classification methods, for which generative AI models are employed to enable deeper and context-sensitive categorization. Furthermore, custom plug-ins are developed to allow users to directly upload and process updated datasets, thereby extending the system's adaptability and inclusiveness. These developments culminate in a more comprehensive and user-driven web-based prototype that enhances interaction with the encyclopedia interface, transforming the abstract taxonomy into a navigable knowledge map. The results are visualized using a variety of hierarchical and interactive diagrams created with advanced data visualization libraries. These include zoomable sunburst charts, radial trees, hierarchical edge bundling, and tree-of-life representations. These visual tools facilitate intuitive navigation through complex knowledge layers, supporting exploratory learning in line with constructivist theory. To evaluate the effectiveness of the system, user experience assessments were conducted through surveys, interviews, and eye-tracking analyses with three distinct participant groups. The results revealed diverse interaction behaviors and gaze patterns, which were used to refine the interface design and enhance overall usability. In conclusion, this study demonstrates that the systematic taxonomy of knowledge into distinct types and

the visualization of their interrelationships can significantly enhance exploratory learning. It presents a novel educational tool that reflects the relational structure of digital architecture knowledge and introduces a new paradigms for its exploration and comprehension.



DİJİTAL MİMARLIK BİLGİSİNİN TAKSONOMİSİ VE GÖRSELLEŞTİRİLMESİ: BİLİMSEL BİR ÇEVİRİM İÇİ ANSİKLOPEDI ÖNERİSİ

ÖZET

Bilişim ve iletişim teknolojilerindeki gelişmeler, insanın bilgiyle kurduğu ilişkiyi dönüştürmüş ve bu dönüşüm, çağın dinamiklerini derinden etkilemiştir. İçinde bulunduğumuz çağa dair “Bilgi Çağı”, “Bilişim Çağı”, “Enformasyon Çağı”, “Dijital Çağ”, “İnternet Çağı”, “Ağ Çağı/Toplumu” gibi bilgiye erişimde paradigma değişikliğini yansıtan çeşitli isimlendirmeler yapılmıştır. Veri trafiğinde yaşanan artışla birlikte verilerden bilgi üretimi, üretilen bilgileri anlamlı hale getirebilme ve bunlardan da yeni bilgi üretimlerini sağlamak amacıyla “Büyük Veri” kavramı ve “Veri Bilimi”, “Veri Madenciliği”, “Veri Görselleştirme” gibi alanlar doğmuştur. Bu alanlarda verinin anlamlı, işlenebilir ve analiz edilebilir bir formata getirilmesi üzerine çalışılır. Böylece veri, artık sadece birikmiş bilgi yığını olmaktan çıkıp, analiz edilmiş, işlenmiş ve anlamlı sonuçların çıkarıldığı bir kaynağa dönüşür. Bilgi, evrimini ve gelişimini sürdürdükçe, söz konusu geniş, katmanlı ve birbirleriyle ilişkili bu bilgileri öğretme ve yayma konusunda yenilikçi yaklaşımları keşfetmek giderek daha da önem kazanmaktadır.

Bilginin üretim ve yayılım biçimindeki bu dönüşüm, bilgi öğretme ve aktarma yöntemlerine yönelik yenilikçi yaklaşımların önemini artırmıştır. Özellikle mimarlık eğitiminde, etkileşimli öğrenme süreçleriyle dijital teknolojilerin entegrasyonu, geleneksel öğretim yöntemlerinin ötesine geçilmesini zorunlu kılmıştır. Dijital teknolojilerin gelişimi, bilimsel bilginin metin ve görseller yoluyla daha etkili aktarılmasına olanak tanımış; bu durum, “dijital mimarlık” gibi yeni alt alanlarda bilgi öğretiminde yenilikçi yöntemlerin kullanımını ön plana çıkarmıştır.

Bu çalışma, kapsamlı ve keşifsel bir öğrenme deneyimi sunmayı amaçlayan bir çevrim içi ansiklopedi aracılığıyla, dijital beşeri bilimlerde yaygın olarak kullanılan metin madenciliği metodunu kullanarak dijital mimarlık bilgisinin hiyerarşik bir şekilde haritalandırılmasına odaklanmakta ve geleneksel, doğrusal bilgi aktarım modellerinin ötesine geçerek, Yapılandırmacılık (Constructivism) ve Keşifsel Öğrenme (Exploratory Learning) teorilerinden beslenen pedagojik bir çerçeve sunmaktadır. Araştırma, dijital mimarlık alanına ilişkin makalelerden oluşan bir veri setini kullanarak, bilgiyi kavramsal, işlemsel, görsel ve konfigürasyonel katmanlar üzerinden yapay zekâ destekli veri madenciliği teknikleri de kullanılarak hiyerarşik olarak sınıflandırmaktadır. Böylece doğrusal metin aktarımının ötesine geçilerek, bilginin katmanlı ve ilişkisel yapısı keşfedici bir yaklaşımla ortaya konmuştur. Bu analitik yaklaşım, dijital mimarlık alanındaki karmaşık bilgi ağının kapsamlı bir şekilde anlaşılmasını sağlar. Elde edilen bulgular, soyut bir taksonominin somut ve gezilebilir bir bilgi haritasına dönüştürülmesini sağlayan etkileşimli görselleştirme teknikleriyle bir web tabanlı arayüz üzerinden sunulmuştur. Bu görsel temsil yöntemleri, kullanıcıların bilgiyi keşifsel bir şekilde öğrenmelerini teşvik eder ve bilgi aktarımını daha etkili hale getirir. Çalışma, dijital mimarlık alanında bilgi

aktarımı ve anlayışını geliştirmeye katkıda bulunmayı hedeflemekte; araştırmacılar, öğretmenler, öğrenciler ve uygulayıcılar için hem görsel bir başvuru kaynağı hem de düşünsel bir araç işlevi görmeyi amaçlamaktadır.

Geleneksel olarak, mimarlık eğitimi dersler, stüdyolar, maketler ve ders kitapları gibi araçları içeren fiziksel sınıf odaklı öğretim yöntemlerine dayanmaktadır. Bu yöntemler birtakım avantajlar sunsa da, ortaya çıkan ve sürekli evrilen dijital teknolojilerin potansiyeli kullanılamamakta ve etkileşimli öğrenme teknikleri eğitim sürecine yeterince entegre edilememektedir. Bu entegrasyon eksikliği, dijital teknolojilerin gelişmesiyle birlikte gelişen, büyüyen bir alan olan dijital mimarlık alanının geleneksel mimarlık eğitimi içerisinde yeterince alan bulamamasına neden olmakta, bu alanın getirebileceği potansiyel profesyonel pratiklerin ve teorilerin de gelişimine etki etmektedir. Dijital mimarlık; mimari süreçlerin hesaplamalı, bilişimsel düşünce ve dijital teknolojiler aracılığıyla kökten dönüştürüldüğü disiplinlerarası bir alandır. Bu yaklaşım, mimarlığı statik bir nesnenin temsili olmaktan çıkarıp, parametreler, algoritmalar ve veri akışlarıyla yönetilen dinamik ve bilgi yoğun bir sisteme dönüştürür. Temelinde yalnızca Bilgisayar Destekli Tasarım (CAD) ve Üretim (CAM) değil, aynı zamanda Parametrik Tasarım, Yapı Bilgi Modellemesi (BIM), performans simülasyonları, jeneratif (üretken) sistemler ve yapay zekâ gibi yeni mimari düşünce paradigmaları yatar. Dolayısıyla bu alandaki kavram, yöntem ve çalışmaların anlaşılması, eğitimdeki entegrasyon eksikliğinin giderilmesi açısından büyük önem taşımaktadır.

Dijital mimarlığın sürekli gelişen, disiplinlerarası doğası ve çok boyutlu yapısı, alana yeni girenler için kavramsal bir karmaşa yaratabilmektedir. Bilginin dağınık ve hızla güncellenen yapısı, öğrencilerin konular arasındaki bağlantılar kurmasını ve edindikleri bilgiyi bütüncül bir çerçevede yapılandırmasını zorlaştırmaktadır. Geleneksel, doğrusal eğitim materyalleri, bu dinamik alanın ilişkisel yapısını ve karmaşıklığını yansıtmakta yetersiz kalmaktadır. Öğrencilerin, öğrendikleri bir konunun genel resimdeki yerini, diğer alanlarla ilişkisini ve hiyerarşik önemini en başından öngörebilmeleri, daha derin ve kalıcı bir anlayış için kritik öneme sahiptir. Bu çalışma, söz konusu eksikliği gidermek amacıyla, dijital mimarlık bilgisini haritalayan dinamik ve etkileşimli bir çevrimiçi ansiklopedi önermektedir. Böylece statik bir basılı kaynak yerine kullanıcıların da eklemeler yapabileceği, kolektif, alanla ilişkili kullanıcılar tarafından erişilebilen üretken bir bilgi ekosistemi oluşturulabilecektir. Böyle bir ekosistem, dijital mimarlık eğitiminde alana ilişkin genel bir perspektifin sunulması ve dijital mimarlık bilgisinin haritalanması, alandaki bilginin kapsamının anlaşılmasının yanı sıra, sonradan edinilecek bilgilerin hangi bağlamda ele alındığına, hangilerinin hiyerarşik anlamda daha önceliklendirilmesi gerektiğine dair bir ön bilginin edinimi açısından da yarar sağlayacaktır. Bu araç, alandaki mevcut bilgi birikiminin bütüncül bir haritasını sunarak yalnızca mevcut durumu anlamayı kolaylaştırmaz; aynı zamanda bilgi boşluklarını tespit etmeye ve yeni araştırma fikirleri geliştirmeye de zemin hazırlar. Geleneksel eğitim yöntemlerinin sınırlarını aşan, keşif odaklı bir çevrimiçi ansiklopedi aracılığıyla dijital mimarlık bilgisini haritalama yöntemi, dinamik ve kapsayıcı bir öğrenme ortamı sunarak bu eksikliği gidermeyi hedeflemektedir. Buna ek olarak, bu çalışma, dijital mimarlık bilgisine dair genel bir bakışın oluşturulmasına ilişkin, kavramsal, işlemsel, görsel ve konfigürasyonel bilgi katmanları üzerinden bir sınıflandırmayı baz alarak, çevrim içi ansiklopedi ile alandaki bilgilerin ilişkiselliklerinin kurgulanarak bilgilerin anlamlandırılmasını, verimli bir şekilde edinimini ve akılda kalıcılığını artırmayı amaçlar. Bu bağlamda çalışma, bilginin doğasına dair temel bir

soruyu da gündeme getirir: Bilgi, lineer olarak tüketilen bir ürün müdür, yoksa ilişkisel olarak keşfedilen bir peyzaj mıdır? Bu tez, ikinci önermeyi savunarak, bu peyzajı gezilebilir kılacak bir arayüz ve metodoloji önerir.

Bu araştırmanın temel çıkış noktasını, bilginin temsili ve edinimi üzerine kurulu iki ana sorunsal oluşturmaktadır:

- 1. Doğrusal metinlerin sunduğu sıralı bilgi akışının aksine, anahtar kelimeler, görseller ve kavramsal bağlantılarla zenginleştirilmiş hiyerarşik bir yapı, bilgi edinimini daha derin ve çağrışımsal bir boyuta taşıyabilir mi?*
- 2. Bilginin sistematik bir şekilde sınıflandırılması ve ilişkisel ağlar aracılığıyla temsil edilmesi, bilginin erişilebilirliğini, anlaşılabilirliğini ve işlevselliğini artırmada ne gibi bir potansiyel taşır?*

Bu ana sorulara yanıt aramak amacıyla, araştırma şu alt sorulara da odaklanmıştır:

- 1. Dijital mimarlık alanına özgü bilginin taksonomik sınıflandırması ve ilişkisel haritalanması için hangi sınıflandırma ve ilişkilendirme olanakları mevcuttur?*
- 2. Dijital mimarlık bilgisinin bütünden parçaya doğru ayrıştırılması, alanın bütüncül olarak kavranmasını nasıl destekler ve bu süreci kolaylaştırmak için hangi görsel temsil araçları etkili bir şekilde kullanılabilir?*

Çalışma kapsamında, dijital mimarlık bilgisinin kavramsal, işlemsel, görsel ve konfigürasyonel bilgi olmak üzere dört bilgi katmanı üzerinden keşif yoluyla ve deneyimsel olarak öğrenilmesine ilişkin hiyerarşik temsil teknikleri kullanılmıştır. Bunun için temel veri tabanı olarak dijital mimarlık alanındaki öncü sempozyumlarda sunulan bildirilerin ve makalelerin bir arada bulunduğu CUMINCAD veri tabanı çalışma kapsamında kullanılmıştır. Bu veri tabanında; ACADIA, ASCAAD, CAADRIA, eCAADe, SiGraDi, CAAD Futures, DDSS gibi önemli akademik platformlardan gelen zengin bibliyografik verileri içeren kapsamlı bir veri havuzu yer almaktadır. Bu çalışma kapsamında CUMINCAD platformundan çekilen makaleler en az 100 kelimelik özetler içermektedir. Bu platform aynı zamanda farklı kaynaklardan veri çekmesi, geniş bir veri setinin bulunması, dijital mimarlık bilgisinin kategorize edilmesini ve .ris, .bibtex, endnote formatlarında verilerin indirilebilmesini sağlaması açısından önemli bir kaynak olma özelliğine sahiptir.

Çalışmanın metodolojisi, iki temel aşamadan oluşan bir veri işleme sürecine dayanmaktadır. İlk aşamada, CUMINCAD veritabanından elde edilen bibliyografik kayıtlar ve kapsamlı özetler, bir dizi Doğal Dil İşleme (NLP) tekniği ile analiz edilmiştir. Bu süreç öncelikle, metin temizleme, TF-IDF vektörleştirici ile terimlerin istatistiksel ağırlığının belirlenmesi ve hiyerarşik kümeleme adımlarını içermiştir. Bu algoritmik işlemler, veri seti içerisindeki örtük tematik yapıları ve temel kavramsal ilişkileri ortaya çıkarmak için temel bir zemin oluşturmuştur. İkinci aşamada ise, ilk aşamada elde edilen bu yapısal bulguları zenginleştirmek amacıyla üretken yapay zekâ modelleri devreye sokulmuştur. Bu modeller, statik algoritmik sınıflandırmanın ötesine geçerek, verinin daha bağlamsal ve anlamsal olarak zengin bir biçimde kategorize edilmesini sağlamıştır. Bu hibrit yaklaşım, yalnızca makalelerin genel örüntülerine göre sınıflandırılmasıyla kalmamış; aynı zamanda her bir makalenin içeriği de kavramsal, işlemsel, görsel ve konfigürasyonel bilgi katmanları ekseninde ayrıca tasnif edilmiştir. Bu çok katmanlı sınıflandırma, geliştirilen çevrim içi ansiklopedi arayüzünde kullanıcının dijital mimarlık bilgisini daha odaklı ve derinlemesine keşfetmesine olanak tanır. Bütünleşik bir metodolojinin kullanılması

ile, dağınık görünen akademik yayınlar arasındaki gizli ilişkiler ve ana desenler görünür kılınarak alanın bütüncül bir bilgi haritası oluşturulur.

Araştırmanın somut çıktısı, kullanıcıların dijital mimarlık bilgisini etkileşimli olarak keşfedebileceği, genel erişime açık bir çevrim içi prototiptir. Bu prototipin arayüzünde, gelişmiş veri görselleştirme kütüphanelerinden yararlanılarak, bilginin hiyerarşik ve ilişkisel doğasını yansıtan çeşitli temsiller üretilmiştir. Verilerin hiyerarşik yapısı; birinci grup diyagramlar olan yakınlaştırılabilir güneş patlaması diyagramı (zoomable sunburst diagram), yakınlaştırılabilir iç içe geçmiş daire kümeleri (zoomable circle packing), sekanslı güneş patlaması (sequences sunburst) ve ikinci grup diyagramlar olan akor diyagramı (hierarchical edge bundling), radyal ağaç (radial tree) ve yaşam ağacı (tree of life) gibi bir dizi etkileşimli görselleştirme tekniği aracılığıyla temsil edilmektedir. Bu dinamik temsiller, kullanıcıya doğrusal bir metnin sınırlamalarının ötesinde, keşfe dayalı ve bütüncül bir öğrenme deneyimi sunarak, farklı bilgi katmanları arasında geçiş yapma ve kavramsal bağlantıları sezgisel olarak kurma imkanı verir.

Çalışmanın pratik uygulamasında iki aşamalı bir geliştirme süreci izlenmiştir. Birinci aşamada, daha küçük ölçekli bir örnek veri seti üzerinde farklı sınıflandırma metodları ve görselleştirme tekniklerinin etkinliği test edilerek bir fizibilite analizi yapılmıştır. İkinci aşamada ise, bu testlerden elde edilen bulgular ışığında, tam veri setiyle çalışacak özel eklentiler (plug-ins) geliştirilmiş ve nihayetinde genel kullanıma açık bir çevrim içi prototip hayata geçirilmiştir. Bu prototip arayüz, kullanıcıların ansiklopediye etkileşimli olarak erişmesine, farklı bilgi katmanları arasında serbestçe gezinerek kavramlar arası ilişkileri keşfetmesine olanak tanımaktadır.

Modelin yukarıda belirtilenler doğrultusunda geliştirilmesine ek olarak, modelin etkinlik ve kullanılabilirliğinin doğrulanması amacıyla, çok yöntemli bir kullanıcı değerlendirme süreci tasarlanmıştır. Bu süreçte, üç ayrı kullanıcı grubu ile çalışılmıştır. Veri toplama sürecinde nicel ve nitel verilerin birlikte toplanabilmesi için anketler, yarı yapılandırılmış mülakatlar ile göz takip (eye-tracking) analizleri gibi yöntemler bütüncül bir yaklaşımla uygulanmıştır. Bu değerlendirmelerden elde edilen bulgular, katılımcıların arayüzle kurduğu etkileşim stratejilerini ve görsel dikkat odaklarını, bakış desenlerini ortaya koymuştur. Toplanan bu veriler, arayüz elemanlarının yinelemeli olarak iyileştirilmesinde ve kullanıcı deneyiminin optimize edilmesinde doğrudan kullanılmıştır. Bu değerlendirmeler, önerilen dörtlü bilgi sınıflandırmasının sadece teorik bir model olmadığını, aynı zamanda kullanıcılar tarafından algılanabilir ve kullanılabilir bir yapı olduğunu ampirik olarak doğrulamıştır.

Bu doğrultuda, araştırmanın temel bulgusu, dijital mimarlık bilgisinin kavramsal, işlemsel, görsel ve konfigürasyonel olmak üzere dört temel katmanda sınıflandırılması ile bu yapıların etkileşimli görselleştirilmesinin, keşfe dayalı öğrenme süreçlerinin etkinliğini artırdığıdır. Bu kapsamda çalışma, alanın karmaşık bilgi yapısını yansıtan ve geleneksel bilgi edinme yöntemlerini yeniden tanımlayan yenilikçi bir çevrim içi ansiklopedi sunmaktadır. Önerilen bu ansiklopedi, hem akademik araştırmacılar hem de alandaki uygulayıcılar için bütüncül, ilişkisel ve sezgisel bir bilgi keşif ortamı yaratır. Nihai olarak bu araç, dijital mimarlık alanındaki pedagojik yaklaşımları, statik içerik tüketiminden dinamik bilgi keşfine doğru dönüştürme potansiyeli taşımaktadır.

Bu tez, araştırmanın sistematik bir şekilde sunulması amacıyla yedi bölüm halinde yapılandırılmıştır. Birinci bölümde, araştırmanın problemi, amaç ve kapsamı tanımlanarak, temel araştırma soruları ve izlenen metodoloji ortaya konarak çalışmanın temeli aktarılır. İkinci bölümde, konuya ilişkin literatür kapsamlı bir şekilde ele alınarak araştırmanın genel kuramsal çerçevesi ortaya konur. Üçüncü bölüm, önerilen metodolojinin merkezinde yer alan temel kavramları ayrıntılı biçimde ele alarak araştırmanın teorik çerçevesini detaylandırır. Dördüncü bölüm, metodolojinin veri toplama, işleme ve görselleştirme adımlarını içeren pratik uygulamayı ayrıntılı olarak sunar. Beşinci bölüm, geliştirilen prototipin kullanıcı testlerinden elde edilen bulguları, anket, mülakat ve göz takip verilerini analiz ederek değerlendirir. Altıncı bölümde, bu bulgular genel çerçevede tartışılır, çalışmanın sınırlamaları ve gelecekteki araştırma olanakları ele alınır. Araştırmanın son bölümünde tüm araştırmanın genel bir sentezi sunularak nihai bulgular ortaya konur ve çalışmanın alana katkıları değerlendirilir.





1. INTRODUCTION

1.1 Statement of the Problem

In the rapidly evolving field of digital architecture, the structuring of knowledge has become a critical necessity for both research and practice. Digital architecture encompasses a broad spectrum of fields of inquiry including computational design methods, parametric modeling techniques, digital fabrication processes and data-driven decision-making tools. As this corpus of knowledge continues to grow in complexity and volume, structuring it in a systematic and meaningful way is essential to support learning, collaboration, innovation, and critical inquiry.

Through the application of text mining and data-driven methodologies, this research introduces a cohesive learning framework designed to support deeper conceptual understanding and foster critical engagement with complex digital architecture content. In professional practice, it enhances the ability to navigate complex workflows, communicate design logic and document processes systematically.

Digital tools amplify the possibilities of knowledge structuring by enabling the visualization, mapping and semantic linking of information. Interactive platforms such as zoomable diagrams, knowledge graphs, and ontologies can reveal hidden relationships and patterns, making the knowledge landscape of digital architecture more navigable and intelligible. Such structured representations also support knowledge sharing within academic and professional communities, fostering collective learning and sustained innovation.

While the principle of knowledge structuring is clear, its implementation is significantly amplified by digital tools that enable the dynamic visualization, mapping, and semantic linking of information. However, a significant gap persists within current pedagogical models. Traditional architectural education has largely relied on conventional, classroom-based teaching methods involving lectures, textbooks, and static visual aids. While possessing their own merits, these approaches often fail to harness the full potential of emerging digital technologies.

The methodology proposed in this study seeks to bridge this gap by designing and implementing an exploratory online encyclopedia, a dynamic and immersive learning environment engineered to transcend the limitations of traditional educational formats.

The objective of this study is to design and implement a novel approach to teaching digital architecture knowledge, encompassing four distinct categories: conceptual knowledge, procedural knowledge, visual knowledge, and configurational knowledge. Conceptual knowledge refers to the theoretical understanding of principles and philosophies in the field of digital architecture. Procedural knowledge involves the practical skills and techniques required to execute digital architecture applications. Visual knowledge relates to the graphical, pictorial and experiential aspects of digital architecture. Configurational knowledge represents the highest level of synthesis, embodying a more integrated and relational approach that connects various components into a coherent whole. Illustrating an instance drawn from the organizational framework of scholarly articles, conceptual knowledge assumes a pivotal role in the introductory section, wherein crucial terms germane to the article are expounded. Procedural knowledge, on the other hand, assumes prominence in the methodological section, delineating the sequential procedures executed in a systematic manner throughout the course of the academic discourse. Confining the scope to visual knowledge, it encapsulates the reproduced visual elements, encompassing graphics, images, tables, among others, embedded within the articles. Configurational knowledge, primarily derived from the results, discussions, or conclusions sections, articulates a comprehensive perspective that predominantly characterizes the overarching essence of the article. Among these types of knowledge, while visual knowledge is typically excluded from abstracts, conceptual, procedural and configurational knowledge are often briefly reflected in the abstract sections of scholarly articles.

To operationalize the categorization of digital architecture knowledge, this study leverages the CUMINCAD database, a repository that serves as the raw dataset for an extensive text mining process. Employing techniques such as topic modeling and advanced clustering algorithms, the research systematically extracts and maps the underlying thematic structures, conceptual relationships, and latent patterns embedded within this scholarly discourse. The resulting structured data is then

rendered through an exploratory visual model, creating a virtual landscape that allows users to navigate the interconnected nature of digital architecture concepts and principles, moving beyond linear text to engage with knowledge spatially and relationally.

The exploratory online encyclopedia serves as the primary platform for delivering this interactive learning experience. Its core navigational logic is structured by hierarchical clustering, which organizes the vast corpus of information into a coherent, step-by-step framework. This structured approach is designed to guide users through the knowledge landscape in an oriented manner, fostering deeper cognitive engagement and enhancing information retention.

The true innovation of this model, however, resides in its layered architecture that integrates multiple dimensions of knowledge representation. The framework of conceptual, procedural, visual, and configurational knowledge is integrated directly into the hierarchical structure, enabling a uniquely focused and comprehensive exploration. Users are thereby empowered not only to navigate between topics but also to deconstruct any given topic through these four distinct analytical lenses. On this basis, this study maps digital architecture knowledge to provide a learning experience that is both immersive and memorable. By leveraging text mining and data-driven structuring, the research delivers an integrated pedagogical model that fundamentally enhances the acquisition and comprehension of complex information. Its primary contribution is a methodology that allows learners to engage with digital architecture knowledge across its conceptual, procedural, visual, and configurational dimensions simultaneously and hierarchically.

While existing text mining applications in scientific research often focus on trend analysis or citation networks, and emerging AI tools like Consensus, ScholarGPT, Google NotebookLM or Scopus AI primarily offer semantic search, content summarization, and reference-based question answering, a significant gap remains in creating a pedagogical framework that structures knowledge for exploratory learning. This thesis addresses this gap by proposing a novel, multi-layered classification model (conceptual, procedural, visual and configurational) specifically designed for the complex domain of digital architecture. Unlike tools that provide answers, this work develops an interactive environment that facilitates question-forming and relational understanding. Its primary contribution, directly reflecting the thesis title

"Taxonomy and Visualization of Digital Architecture Knowledge," is therefore twofold: first, the development of a robust taxonomy through its classification model, and second, the dynamic visualization of that structure. This visualization transforms the abstract classification into an interactive and navigable knowledge map. It is this visual and explorable structure that supports a constructivist approach to learning, transforming the user from a passive information consumer into an active explorer of the disciplinary landscape.

1.2 Purpose, Scope and Method of the Thesis

The following goals will be pursued by the thesis study as it explores the effects of integrating visualization tools into the instruction of digital architecture terms:

-The goal of the thesis is to determine how visualizations may help people grasp complicated ideas in digital architecture in a more straightforward and natural way. The usefulness of charts, diagrams, and other visual aids in assisting learners in understanding complex or abstract concepts is examined. The underlying pedagogical objective is to create a constructivist learning environment that promotes exploratory learning (Bruner, 1961). By enabling users to navigate knowledge non-linearly and discover relationships for themselves, the system aims to foster deep learning and the development of robust relational models, rather than simple information transmission.

-The study investigates how visualizations might improve learning interaction and engagement. It attempts to assess how dynamic visualizations, like interactive graphs, promoting a more engaged and active learning environment.

-The purpose of the study is to determine whether or not students can retain important terminology related to digital architecture better when they are presented with visual aids in addition to textual material.

The scope of this study is to explore the mapping of digital architecture knowledge using an exploratory online encyclopedia. The study focuses on teaching knowledge in four categories: conceptual, procedural, visual, and configurational (Figure 1.1). The research methodology involves scraping data from the CUMINCAD database and employing text mining techniques such as topic modeling and clustering algorithms to uncover relationships between different concepts and to classify

hierarchically according to different topics. The data downloaded from CUMINCAD database contains a text file in .ris format includes the titles of the articles, the keywords of the articles, the summaries of the articles, the authors, the publication years of the articles and the conferences' names. The titles and the keywords of the articles are used in the study primarily. At the initial stage of the study, the text file was converted into an Excel file containing only the titles and keywords of the articles. The dataset originally comprised 9,808 articles, which was subsequently reduced to approximately 6,000 after text preprocessing, retaining only those entries with keyword information.

This study offers a novel contribution by bridging two traditionally siloed fields of inquiry through an integrative framework: the deep hierarchical classification of scholarly knowledge and its interactive, multi-modal visualization. While the literature contains numerous examples of either advanced text classification or data visualization, few approaches integrate both into a single, cohesive framework designed for rapid and intuitive knowledge exploration. To bridge this gap, the research proposes the development of an exploratory data model built with the D3.js library. The implementation leveraged the capabilities of generative AI¹ to develop the complex structural code necessary for constructing the visual vocabulary. This includes a range of hierarchical diagrams such as zoomable sunbursts, circle packing, sequences sunbursts, radial trees, tree of life representations, and hierarchical edge bundling, all designed to make intricate knowledge structures visually intelligible. Following its initial development, the web-based prototype was subjected to a preliminary evaluation by a student cohort, with the gathered qualitative and quantitative feedback being used to systematically refine its interface and user experience.

¹ In this study, generative AI tools, such as ChatGPT and Google AI Studio, were utilized in coding applications.

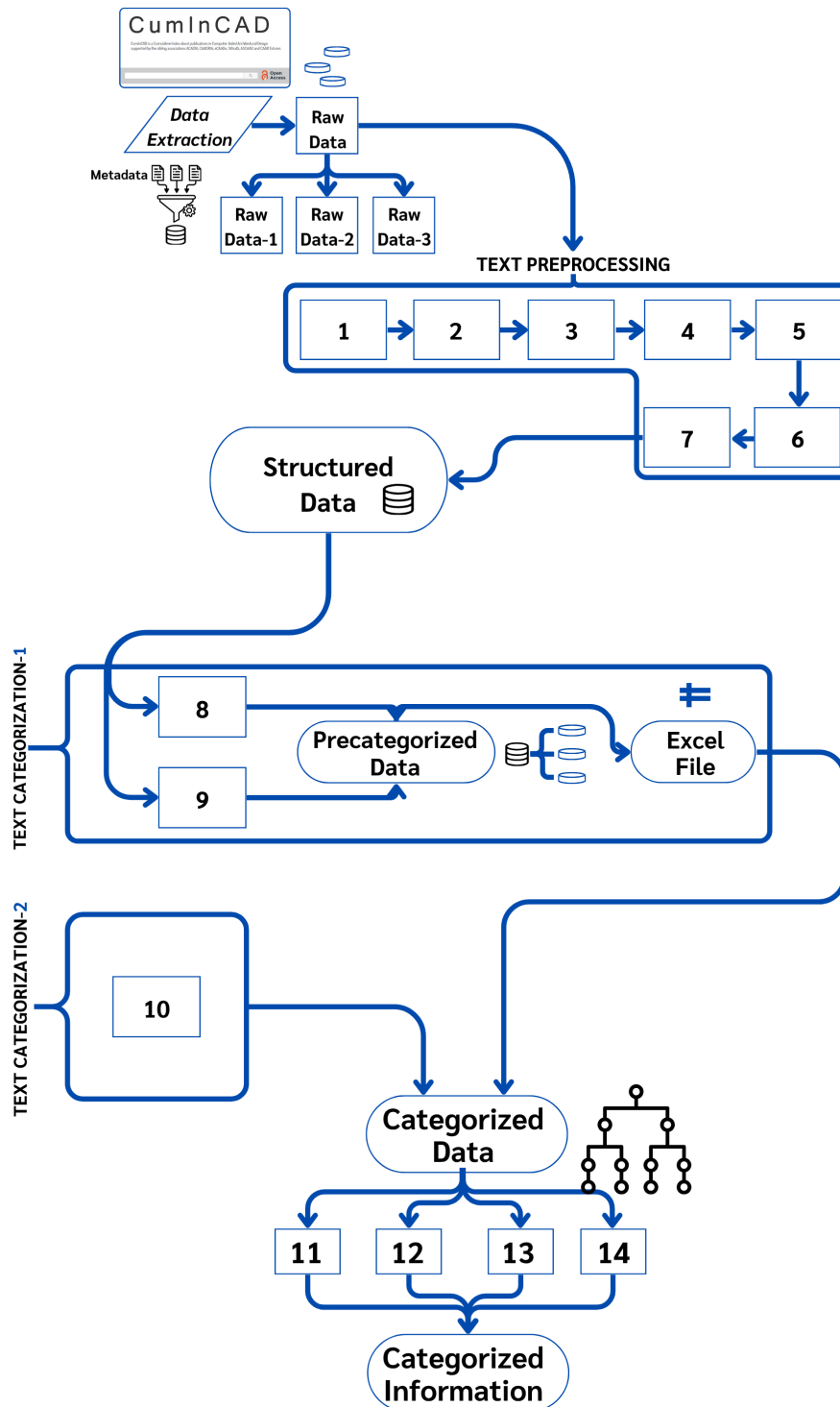


Figure 1.1 : The flowchart of the methodology.

The explanations of each step is as follows (Figure 1.1):

1. Removal of Rows with Cells Containing Non-English Words
2. Removal of Columns with Keywords Containing Non-English Words
3. Capitalization of Each Word

4. Correction of Spelling Mistakes
5. Removal of Rows with Empty Cells
6. Extraction of File with Separated Keywords and Removal of Cells with More Than 5 Words
7. Lowercasing of Certain Common Words, Excluding Those Found at the Beginning of a Title (of, on, and, to, a, in, for, by)
8. Conversion of texts into numerical data with TF-IDF text vectorization method
9. Classification of articles according to their titles and keywords using k-Means clustering algorithm
10. Creation of RIS to JSON Converter and Categorized Hierarchical Keyword Classification Plug-Ins
11. Conceptual Knowledge
12. Procedural Knowledge
13. Visual Knowledge
14. Configurational Knowledge

This procedure outlines a systematic approach to preparing and analyzing textual data for classification. Initially, rows with non-English words and columns with keywords containing non-English terms are removed to ensure linguistic consistency. Text is standardized by capitalizing each word, correcting spelling errors, and eliminating rows with empty cells. Keywords are extracted into a separate file, with entries exceeding five words removed for conciseness. Common words like "of," "on," and "to" are lowercased unless they appear at the beginning of a title. The dataset is then converted into numerical form using the TF-IDF (Term Frequency-Inverse Document Frequency) vectorization method, which quantifies the importance of terms (Salton and Buckley, 1988). Articles are grouped based on their titles and keywords using the k-Means clustering algorithm, which identifies patterns and organizes data into clusters in the first phase of the methodology (Duda and Hart, 1973; Hartigan, 1975).

The study is structured in two main phases: the first involves the testing of classification and visualization processes on a sampled dataset, while the second

phase expands the scope by employing a full dataset, incorporating the development of plug-ins and a web-based prototype for public interaction. A set of user experience evaluations, including surveys, interviews, quizzes and eye-tracking tests, were conducted across three participant groups. These analyses reveal patterns in attention distribution and user behavior, contributing to the refinement of interface design and visualization strategies. Findings suggest that structuring knowledge types contributes significantly to exploratory learning, offering users a deeper and more organized understanding of the complex relationships in digital architecture. The study bridges methods from digital humanities, educational technologies, and architectural computing, proposing a novel model that not only visualizes knowledge but also redefines how it can be accessed, navigated, and learned. Finally, articles are classified into four knowledge categories; conceptual, procedural, visual, and configurational, based on their content, ensuring a clear organization of information for analysis.

2. EXPLORING THE IMPACT OF KNOWLEDGE STRUCTURING

2.1 Effects of Knowledge Categorization on Learning

The structuring of information into meaningful categories is a foundational cognitive process, essential for both knowledge acquisition and basic perception. As Goldstein (2008) posits, without the cognitive shortcut of categorization, our perception of the world would be overwhelmingly complex, forcing us to learn about every object as a unique, isolated entity. This principle is echoed in instructional design theory, where Sweller et al. (2011) identify the categorization of knowledge as a primary step in organizing human cognition. A central challenge for educators, therefore, is to understand how different types of knowledge are processed and retained, as this dictates the most effective instructional strategies.

Despite its fundamental role, the discourse surrounding this process is fragmented by terminological inconsistencies. Interdisciplinary research is challenged by the overlapping use of closely related terms like classification, conceptualization, taxonomy, and typology, which are often employed interchangeably despite nuanced distinctions. A notable distinction, highlighted by Tonta (2012), exists between "classification," which is prevalent in librarianship and information science, and "categorization," which is more commonly used in computer engineering and cognitive sciences. This terminological variance underscores the need for clarity when discussing knowledge organization.

Beyond terminology, the mechanisms of categorization themselves are multifaceted. The most straightforward is the definitional approach, where an object's membership is determined by whether it meets a set of fixed criteria. However, this method proves insufficient for many human-made objects; for instance, the dictionary definition of a "chair" fails to encompass the vast diversity of contemporary designs (Oxford Learner's Dictionaries, 2024). To address these limitations, cognitive science proposes more flexible models such as family resemblance, prototype approach (finding the average case), and exemplar approach, exist in relation to categorization.

Amadiou et al. (2009) conducted an investigation into the impact of prior knowledge on cognitive load during the acquisition of knowledge from non-linear hypertext concept maps within the field of biology. The structured maps explicitly illustrated the hierarchical relationships among concepts, while the unstructured maps portrayed only the network of relations without a distinct hierarchy. The findings revealed that learners with low prior knowledge exhibited greater gains in conceptual knowledge when exposed to the structured format. Moreover, these low prior knowledge learners reported experiencing less cognitive load in the posttest performance following exposure to the structured format. Conversely, high prior knowledge learners showed no discernible difference in conceptual knowledge gains between the two formats. Both low and high prior knowledge learners indicated reduced cognitive load when processing the structured concept maps. Additionally, Amadiou et al. (2009) established that, for low prior knowledge learners, the hierarchical structure significantly enhanced their free recall of the material. In contrast, high prior knowledge learners demonstrated superior performance after studying the unstructured format. Notably, low prior knowledge learners benefitted from the information presented in the hierarchical structure, whereas the same information proved redundant for high prior knowledge learners. Consequently, processing this redundant information imposed an extraneous cognitive load that hindered further learning, thereby manifesting a classic expertise reversal effect.

The comprehension of knowledge requires moving beyond simple categorization to understand the intricate relationships between concepts. The semantic network approach provides a foundational model for this relational understanding, positing that knowledge is organized in a meaningful, interconnected web. A seminal example is the hierarchical network model proposed by Collins and Quillian (1969). In this model, knowledge is visualized as a network of nodes, where each node represents a concept (e.g., "bird") and the links between them signify semantic associations (e.g., the link between "canary" and "bird") (Figure 2.1).

A key principle of this hierarchical organization is cognitive economy. As described by Goldstein (2008), this principle dictates that properties are stored at the highest possible node in the hierarchy to avoid redundancy. For instance, the property "can fly" is stored at the "bird" node rather than being repeated for every individual bird type. This structure allows for efficient information retrieval through inheritance,

demonstrating how knowledge can be organized for both logical coherence and processing efficiency.

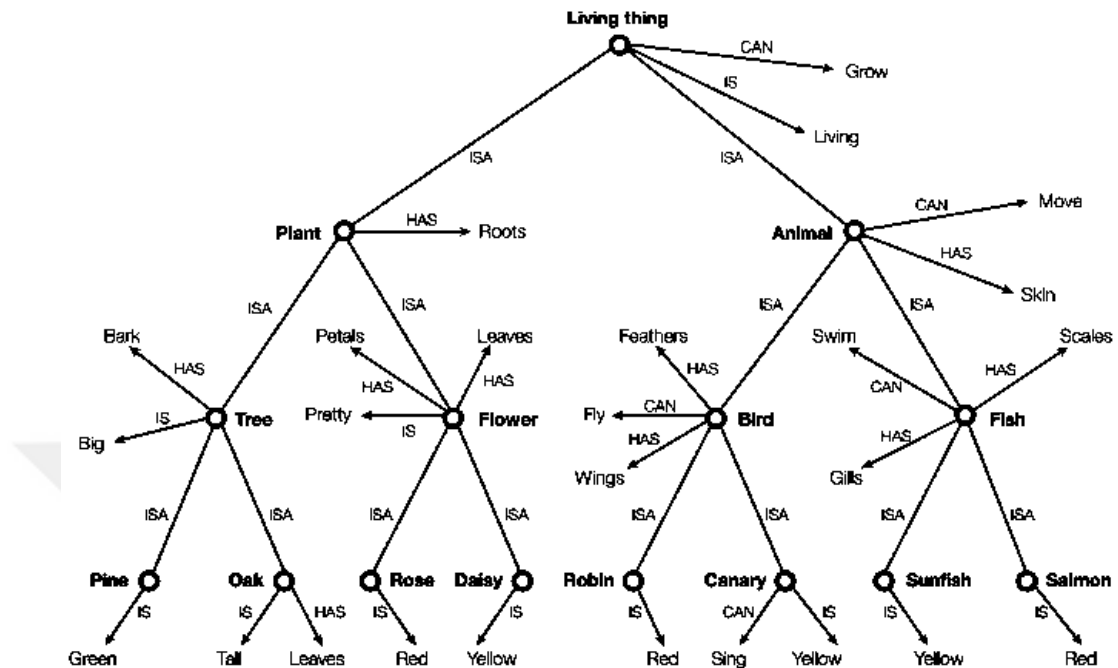


Figure 2.1 : The hierarchical propositional model of Quillian (McClelland and Rogers, 2003).

The connectionist approach offers a distinct computational paradigm inspired by the neural architecture of the brain. Also known as Parallel Distributed Processing (PDP), it posits that concepts are not stored in discrete, symbolic nodes but instead emerge from patterns of distributed activity across a network. In these models, a given stimulus activates a set of 'input units,' and this activation propagates through interconnected layers of 'hidden units' to a final layer of 'output units' as depicted in Figure 2.2. The crucial element is the connection weight, which modulates the strength of the signal between any two units. The cumulative effect of these weights therefore determines the final pattern of activity in the output units, a pattern that represents the network's response. (McClelland & Rogers, 2003). This model forms the theoretical foundation for modern artificial neural networks (ANNs), where these connection weights are dynamically adjusted through a learning process, allowing the network to adapt and recognize new patterns.

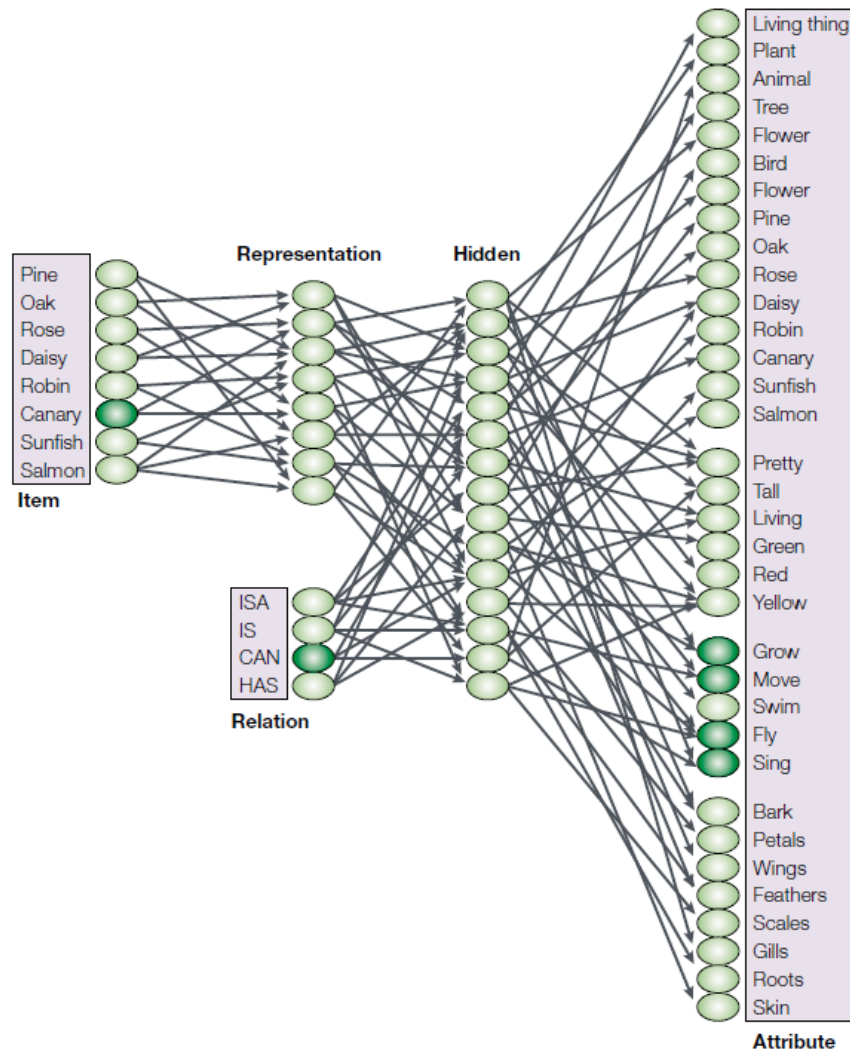


Figure 2.2 : Connectionist model, parallel distributed processing network (McClelland and Rogers, 2003).

2.2 Knowledge Types

The classification of knowledge, the central component of learning, is a foundational concern in educational and cognitive sciences. The literature presents a vast and diverse landscape of knowledge typologies, often categorized by their nature or mode of acquisition. These classifications range from fundamental philosophical distinctions, such as a priori and a posteriori, to cognitive categories like procedural, conceptual, and metacognitive knowledge. Further distinctions include experiential, tacit-explicit, and situational knowledge, among many others detailed in Table 2.1. The sheer breadth of these classifications highlights a critical point: no single type of knowledge is sufficient to fully describe a complex phenomenon. Instead, these

different forms of knowledge are often invoked in concert, creating a composite understanding where one type complements the limitations of another.

Table 2.1 : Classification of knowledge types according to different authors.

Author Names	Knowledge Types
Bradley and Swartz, 1979	Experiential, Ratiocinative, Empirical, A Priori
Reif, 1987	Declarative, Procedural, Formal, Compiled, Special, General, Procedural Interpretation, Coherent
Alexander and Judy, 1988	Declarative, Procedural, Conditional
Kogut and Zander, 1992	Explicit, Implicit
Reif and Allen, 1992	Main Interpretation, General, Definitional, Ancillary, Supplementary, Case-Specific, Entailed, Concept
Jonassen et al., 1993	Structural
Jong and Ferguson-Hessler, 1996	Situational, Conceptual, Procedural, Strategic
McCormick, 1997	Conceptual, Procedural
Eppler and Burkhard, 2004	Declarative, Procedural, Experiential, Orientational, Individual
Asheim and Coenen, 2005	Analytical, Synthetic, Symbolic
Blumberg, 2009	Factual, Conceptual, Procedural, Meta-cognitive
Selvi, 2011	Phenomenological
Millar, 2019	Perceptual
Russell, 2020	A Priori, A Posteriori
Vercruysse and Kuiper, 2020	Computable
Lynch, 2022	Symbolic, Intuitive
Zhao et al., 2024	Analytical, Symbolic

When examining types of knowledge, no universally accepted classification exists among scientists. However, systematically addressing “types of knowledge” can facilitate the creation of a layered, hierarchical structure for comprehensive acquisition of concepts. This approach can also help in the gradual transmission of knowledge by reducing cognitive load (Sweller et al., 2011; Amadiou et al., 2009).

Among various knowledge types, conceptual and procedural knowledge are most commonly included in classifications. McCormick (1997) describes procedural knowledge as "knowing how to do" something and conceptual knowledge as the ability to relate different parts. Anderson (1976) distinguishes procedural knowledge, knowing "how" to do something, from declarative knowledge, knowing "what" something is. Building upon this foundational distinction between types of knowledge, subsequent studies have focused on how these knowledge forms can be systematically categorized and leveraged within design processes. Goldstein (2008) states that the most studied function of concepts is their role in categorizing entities into groups known as categories. In the computational design field, Koutamanis (2000) explores the categorization into visual and conceptual components, stating that "visual representations play a crucial role in the structuring of design knowledge," directly relating to structured information in architectural design. Similarly, Gero and Maher (1993) provide a model of knowledge-based creative design, stating that "knowledge categorization can influence creative design processes by providing a structured framework for understanding design problems and generating solutions," offering a framework for understanding the influence of categorized knowledge in design processes and outcomes.

Research in the field of science education on visual knowledge or the visual representation of knowledge is generally related to cognitive load theory. According to Cook (2006), although presentations combining visual and verbal information are widely used to display instructional materials in modern science education, educational research has been slow in determining how to design these learning materials effectively. Cook (2006) emphasizes that understanding the cognitive foundations that enable the effective representation of scientific concepts is crucial for creating instructional materials that utilize multiple representation forms. The studies of Sweller et al. (2011), based on cognitive load theory, have resulted in empirical evidence supporting design principles derived from this theory. For educational materials to contribute to students' knowledge acquisition, their visual design must also be considered. Visual representations function as more than mere illustrations; they are fundamental cognitive tools that translate complex, abstract, or multi-dimensional scientific data into perceptually intuitive forms. By doing so, they significantly reduce cognitive load, facilitating a more rapid comprehension of latent

relationships and structural patterns, and thereby enabling a deeper and more robust understanding of the core concepts. In addition to these three types of knowledge, configurational knowledge is not defined in the fields of education, cognitive psychology, knowledge management, and epistemology. However, in fields like environmental psychology and architecture, configurational knowledge can be described as the connection of relationships and transitions within a space. Architectural and urban design research investigates 'how things are put together' and how to reveal complex patterns (Hillier, 2007). Jonassen et al. (1993) describe structural knowledge, which is akin to configurational knowledge, as the understanding of how concepts are interrelated. This study emphasizes configurational knowledge as the arrangement and relationships within structures. The research highlights the visualization of four types of knowledge (conceptual, procedural, visual, and configurational) and their interdependence, enabling comprehensive information acquisition through exploration.

The scope of the research topic encompasses scientific metadata in the field of digital architecture. In this context, conceptual knowledge, procedural knowledge and configurational knowledge play significant roles in scientific studies. Conceptual knowledge aids in understanding various concepts within the research scope and the relationships among them, while procedural knowledge details the methodological steps of scientific studies. This type of knowledge explains step-by-step how the research will be conducted. Configurational knowledge conveys the relationships among the fundamental information contained in scientific studies. Within this framework, conceptual knowledge involves the hierarchical mapping of concepts in articles; procedural knowledge represents the flowchart of procedural steps; and configurational knowledge is represented through knowledge network maps. Although visual knowledge is not present in all scientific studies, it has been incorporated in this study through the representation of four distinct visual types (render, sketch, technical drawing, and pictogram) generated using Microsoft Designer. Microsoft Designer is an AI-supported design platform that facilitates the creation of visual content by integrating machine learning algorithms and generative models, thereby streamlining design processes and enhancing user efficiency in producing professional-quality outputs. The representation of visual knowledge is important for the perceptual and experiential acquisition of information. Thus, the

interaction between different types of knowledge and their understandability are enhanced. This approach, supported by cognitive load theory, helps optimize students' learning processes (Cook, 2006; Sweller et al., 2011).

2.2.1 Conceptual knowledge

Conceptual knowledge constitutes a higher-order cognitive framework that systematically organizes abstract principles, theoretical constructs, and mental models into a coherent structure. It transcends a superficial awareness of discrete facts by addressing the underlying "why" and "how" of phenomena, thereby enabling deep understanding that is both meaningful and transferable across contexts. While perceptual knowledge, akin to Kant's notion of a posteriori, derives from immediate sensory experiences and provides raw, unprocessed data, conceptual knowledge entails the active mental structuring of such data into stable, generalizable systems of meaning.

The development of conceptual knowledge marks a pivotal cognitive transition from passive reception to active interpretation. Solomon et al. (1999) refer to this as the construction of a "mental architecture" that underpins complex intellectual functions such as reasoning, memory, and language processing. This mental architecture typically encompasses both declarative and propositional knowledge, which involves the articulation of factual and conditional information. Schoenfeld (1985) underscores the significance of this transformation, asserting that procedural knowledge, when detached from an underlying conceptual base, tends to be "dry and focused on rote memorization." Conceptual understanding, in line with this, imbues procedures and factual information with meaning and intentionality.

A key domain where conceptual understanding demonstrates its cognitive utility is categorization. As Goldstein (2008) explains, human cognition continuously relies on the classification of objects and experiences into categories, for example, identifying and differentiating between "passenger cars" and "minibuses." This process is not merely taxonomic; rather, it enables individuals to infer properties, predict behaviors, and allocate cognitive resources efficiently. The ability to categorize thus represents a fundamental mechanism of cognitive economy, a view further supported by Solomon et al. (1999).

In architectural discourse and practice, conceptual knowledge plays an indispensable role. Differentiating between architectural paradigms such as “modernism” and “postmodernism” is inherently a conceptual task, necessitating familiarity with abstract design principles, historical trajectories, and philosophical underpinnings. Conceptual knowledge enables students and practitioners alike to interpret architectural works not merely in terms of physical form, but as manifestations of theoretical, cultural, and ideological frameworks. Ultimately, it is this level of cognitive abstraction that fosters critical engagement, reflective design thinking, and disciplinary depth.

2.2.2 Procedural knowledge

Procedural knowledge, one of the types of practice-based knowledge, focuses on knowing how to perform a task and is concerned with practical skills rather than theoretical understanding. McCormick (1997) defines procedural knowledge as the knowledge of "knowing how to do." According to McCormick (1997), procedural knowledge is considered to be in an active form because it involves the ability to perform tasks or actions, rather than just knowing facts passively. This knowledge is enacted through doing, requiring the learner to apply skills and techniques in practical contexts, which distinguishes it from conceptual knowledge that is primarily about understanding and recalling information. In the view of Bumen (2006), procedural knowledge encompasses information related to skills and algorithms, methods, techniques, and criteria. Anderson (1976), in distinguishing between procedural and conceptual knowledge, uses the expressions "knowing how" for procedural knowledge, which involves understanding how to perform a task, and "knowing what" for conceptual knowledge, which involves understanding what something is.

Procedural knowledge is understanding of how to perform something. It has to deal with the know-how method of a transaction. It explains how to use the information and how to perform the tasks step by step. Within this knowledge layer, several forms of knowledge, including imperative knowledge, operational knowledge, practical knowledge, knowledge of techniques, and knowledge of tools, can be evaluated. The application of the information is structured. It is universal and objective since it explains step-by-step how to create an application using specific

tools under specific circumstances, but it also permits fresh and unique productions through interventions with variables.

Procedural knowledge refers to the skills and techniques that are applied in the real world when carrying out tasks. Methodologies, processes, and experimental approaches are frequently mentioned in scientific literature. Numerous scientific articles provide step-by-step instructions for experimental protocols, approaches, and processes. By classifying this information under procedural knowledge, the articles' practical instructions and procedures are acknowledged for their significance.

Examples of procedural knowledge include the ability to use BIM tools for generating building systems or using parametric design workflows. For instance, understanding the sequence of steps required to construct a Revit family or to run an environmental analysis in Grasshopper reflects procedural knowledge.

To visually represent and deconstruct these procedural sequences, this study utilizes interactive flowcharts. Developed using the GoJS library for JavaScript, these diagrams articulate the step-by-step logic of a given process, from data collection to final analysis. A key feature of this implementation is that the flowcharts are not static; users are empowered to directly edit and manipulate the diagrams. This interactive capability allows learners to experiment with the procedural logic, test alternative sequences, and thereby develop a deeper, more hands-on understanding of the operational aspects of digital architecture.

2.2.3 Visual knowledge

Visual knowledge is defined as the active capacity to interpret, analyze, and generate visual-spatial information, a skill that is paramount in disciplines like architecture and design. This form of knowledge transcends mere perception; for example, it involves complex cognitive processes such as reading a technical drawing to understand the spatial relationships between floors or mentally visualizing the transformation of a 2D plan into a 3D structure. While it is rooted in the broader category of perceptual knowledge, which, as Millar (2019; 2000) notes, involves the direct, sensory-based recognition of objects like identifying a birch tree on sight, visual knowledge is a more specialized and active faculty.

The decision to concentrate on visual knowledge in this study is a deliberate and strategic one. Scientific communication, particularly within the CUMINCAD

database, heavily relies on visual artifacts to convey complex ideas. Scholarly articles extensively incorporate graphical elements such as graphs, charts, schematic sketches, technical drawings, renderings, and pictograms, which serve as integral vehicles for conveying complex information, all of which function as critical carriers of meaning. This study recognizes that the ability to decode these representations is essential for a comprehensive understanding of the research presented.

Furthermore, this focus is a pragmatic acknowledgment of the opportunities and limitations presented by the dataset. While a holistic understanding of a phenomenon might involve multiple senses, the textual and graphical nature of scholarly publications makes visual knowledge the most accessible and analyzable modality. This research therefore employs AI-driven methods to systematically extract and categorize these visual elements. By concentrating on this specific type of knowledge, the study adopts a practical approach that maximizes the potential of the available data, aiming to understand how visual information is structured, communicated, and utilized within the discourse of digital architecture.

2.2.4 Configurational knowledge

Configurational knowledge refers to knowledge related to configurations or arrangements of elements within a system or structure. Configurational knowledge encompass an understanding of how individual components are organized, interconnected, or arranged to form a coherent whole. In the given context, "configurational knowledge" is not a widely recognized term, but it holds significance within the architectural field. This term is particularly common in architectural discourse where the analysis of spatial arrangements and configurations plays a crucial role.

In architectural context, the analysis of spatial arrangement is based on the perspective that "socially and culturally defined patterns are concealed within the spatial composition" and encompasses the representation and digital analysis of spatial configuration (Hillier et al., 1987). This technique investigates architectural space through the lens of configurational properties that reflect the social functions indicated by spatial layouts and the cultural significance they embody. To comprehend spatial configuration, it is necessary to examine the proximities arising from the interrelatedness of spaces. According to Hillier (2014), the relationship

configuration between two spaces cannot be determined without a third space because the relationship between two spaces serves only as a simple connection, whereas configuration aids in understanding the holistic relationships.

Configurational knowledge represents the highest level of structured understanding, pertaining to how individual components of a system interrelate to form a coherent and functional whole. It moves beyond the analysis of isolated elements to a holistic comprehension of a system's overarching logic and elemental organization. For instance, in architectural design, this knowledge is exemplified by the ability to grasp how circulation, lighting, and structural systems are intricately interwoven in a complex building like a museum. It is the understanding not of the parts in isolation, but of the integrated synergy that defines the design.

In the context of scientific communication, configurational knowledge extends beyond spatial arrangements to include the formal structure of scholarly articles. The standardized format (introduction, methods, results, and conclusion) reflects an implicit hierarchy that guides interpretation. The placement of results, discussion, and conclusion sections within this configurational framework is particularly significant; it underscores the importance of these sections in synthesizing prior information and communicating the article's core scientific contribution and overarching perspective.

To visually articulate and explore these complex configurational relationships, this study implements an interactive knowledge graph. Developed using the GoJS library for JavaScript, this tool serves as a dynamic representation of the interconnected knowledge structure. By allowing users to manipulate the graph, the platform not only presents configurational knowledge but also enables an active, exploratory engagement with it, fostering a deeper and more integrated understanding of the discipline's holistic structure.

2.3 Representation and Mapping of Knowledge

2.3.1 Visual representation of knowledge

The visual representation of knowledge is the transformative process of externalizing complex, abstract informational structures into perceptible and tractable forms. Its

primary role is to render knowledge cognitively accessible and analytically traceable, enabling both a holistic overview and the detailed exploration of new linkages between classified concepts. In this capacity, visualization serves as a vital instrument for constructing comprehensive mental models.

Functioning not merely as aesthetic expression but as a critical cognitive and epistemological instrument, visualization mediates between abstract thought and tangible articulation. Through established techniques such as mind maps, flow diagrams, hierarchical trees, and network graphs, knowledge is systematically organized and recontextualized. This process, in turn, reveals latent patterns, structural hierarchies, and associative clusters that might remain obscured in purely textual or numerical formats.

This epistemological function is especially vital within architecture, a discipline where reasoning is inherently spatial, visual, and systemic. Here, visual representations bridge the gap between the conceptual and the perceptual, providing graphical mappings that are integral to developing a relational understanding of design concepts, theoretical frameworks, and contextual data. Ultimately, these representations support the formation of dynamic knowledge ecologies, providing a visual framework that integrates the conceptual, procedural and configurational dimensions of the discipline into a cohesive whole.

2.3.1.1 Creative mapping in architecture theory

In the domain of architectural theory, creative mapping has been established as a potent methodological tool. Unlike traditional cartography, which documents existing territories, these maps function as speculative instruments designed to visualize the intricate and often non-linear relationships between ideas, disciplines, and spatial practices. They do not merely represent knowledge; they actively interpret and reorganize theoretical constructs through a distinct visual language.

A comprehensive taxonomy of these approaches is offered by Güleç (2021), who identifies ten distinct categories of creative maps, including mental, theoretical, narrative, figurative, and relational maps, among others. This framework provides a valuable lens for understanding the multiplicity of visual strategies employed in contemporary architectural theories. The diversity of these types underscores the flexibility of mapping as a mode of inquiry.

In the context of creative mapping, Jencks's work serves not merely as a passive record of historical information but as an interpretative visual model that invites critical engagement. It transforms complex theoretical content into an accessible and structured visual language, exemplifying how conceptual knowledge can be externalized through visual representation. Moreover, the map supports both top-down and bottom-up reading strategies, allowing users to trace overarching movements or focus on specific schools of thought, thus enabling both macro and micro-level comprehension.

In a distinct approach, Richard Giblett's 21. Mycelium Rhizome (2009) takes a radically different approach to mapping by employing organic metaphors and nonlinear structures. Inspired by biological systems, particularly fungal networks and Deleuzian rhizomatic theory, this work visualizes knowledge not as a hierarchical or evolutionary construct, but as a dense, interconnected web with no clear origin or end. According to the principles of connection and heterogeneity, any point within a rhizome is capable of being linked to any other, and such connections are essential. This stands in stark contrast to hierarchical structures like the tree or root, which establish fixed positions and impose an ordered, linear arrangement (Deleuze and Guattari, 1987). The rhizome becomes a visual metaphor for decentralized thinking, multiplicity, and intersubjective relations within knowledge systems.

In contrast to Jencks's structured historical mapping, Giblett's piece is speculative, abstract and evocative. It emphasizes the relational and interactional aspects of knowledge, challenging conventional taxonomies and linear progressions. Through its intricate network of interlacing nodes, lines, and connections, the piece illustrates how knowledge can be simultaneously spatial, temporal and conceptual, resisting fixed categorization and inviting continuous reinterpretation (Figure 2.4).

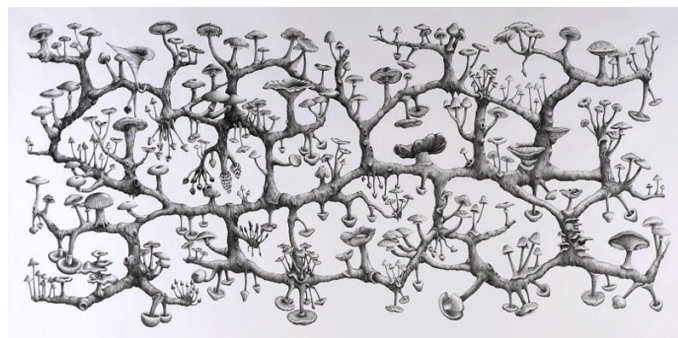


Figure 2.4 : Richard Giblett, 21. mycelium rhizome (Giblett, 2009; Menken, 2019).

Giblett's work encourages viewers to navigate the diagram intuitively, echoing the nonlinearity of contemporary digital knowledge ecologies, such as hypertext systems and online encyclopedias.

Together, Jencks's and Giblett's mappings illustrate two contrasting but complementary approaches to creative mapping in architecture. While Jencks leans toward systematic categorization and historiographic clarity, Giblett foregrounds complexity, ambiguity and emergence. These works demonstrate the epistemological spectrum that visual mapping in architecture can occupy from structured and didactic to open-ended and experiential. Importantly, both examples reinforce the argument that creative visual mapping is not merely illustrative but constitutive of architectural knowledge itself.

The interrelation between knowledge and mapping constitutes a fundamental mechanism through which complex information is conceptualized, structured and communicated. Far beyond its conventional role in geographical representation, mapping functions as an epistemological and cognitive framework that facilitates the organization, interpretation and visualization of knowledge, particularly in spatial, theoretical and conceptual domains. According to Dodge et al. (2011) mapping offers a way to structure extensive and often multidimensional information about a place, making it more accessible for human exploration and comprehension. Polanyi (1963) frequently employs maps and mapping in his exploration of knowledge and the process of knowing. He considers maps a clear example of explicit knowledge, a form of factual representation, which allows for critical examination both by the creator and subsequent users. Maps are especially useful in this context because they can be readily verified, making it easier to reach a consensus on their accuracy compared to other forms of knowledge, such as written texts.

2.3.1.2 Visual-spatial mapping of knowledge

Beyond conventional diagrammatic representations, the spatial dimension itself can act as a medium for knowledge visualization. Visual-spatial mapping combines visual signifiers with the physical organization of space to create immersive, multi-dimensional knowledge environments. This approach transcends flat visual surfaces and incorporates bodily movement, sensory perception, and spatial hierarchy into the process of understanding and communicating knowledge.

One of the historical exemplars of this method is the work of Patrick Geddes, a pioneering figure in urban planning and sociological geography. Geddes conceptualized spatial mapping as an integrative tool for both analysis and education. He designed the spatial layout of his Outlook Tower in Edinburgh to physically manifest a hierarchy of information, from broad, global perspectives on the lower levels to increasingly localized and specialized knowledge as one ascended the tower. This architectural structure functioned as a three-dimensional epistemological diagram: space was not just a container for knowledge but was knowledge itself, organized according to conceptual gradation. Geddes' approach embodies a deep understanding of how spatial organization can reflect and reinforce the structure of knowledge systems (Acar, 2019).

In contemporary contexts, this spatial logic has been extended through the use of interactive media, virtual reality, and digital cartographies. Tools such as GIS, immersive data environments and 3D modeling software allow for the construction of responsive knowledge spaces where users can explore information not only visually but also spatially and experientially.

Visual-spatial mapping thus provides a powerful paradigm for holistic representation, enabling users to navigate through abstract knowledge in concrete, embodied ways. It aligns particularly well with architectural and design-oriented disciplines, where the spatial arrangement of information mirrors the cognitive and procedural logics of the domain itself.

2.3.2 Textual representation of knowledge

Textual representation plays a fundamental role in the conceptualization and communication of knowledge. Following initial perceptual processes, knowledge is often abstracted and refined through language, enabling complex ideas to be shared, debated, and preserved across time and space. Textual knowledge representation tools such as dictionaries, encyclopedias, glossaries, and scholarly texts provide a systematic framework to define, elaborate and contextualize concepts. This form of knowledge representation not only facilitates comprehension but also supports the critical examination and continuous evolution of ideas within and across disciplines. The structured organization of textual knowledge allows users to navigate, compare

and synthesize information effectively, forming a crucial complement to visual and experiential learning modalities.

2.3.2.1 Printed encyclopedias and dictionaries

Printed encyclopedias and dictionaries have historically served as authoritative repositories of knowledge, offering meticulously curated and peer-reviewed content. The *Encyclopedia or a Systematic Dictionary of the Sciences, Arts, and Crafts* by D'Alembert and Diderot (1751) marks a seminal milestone in this tradition. Unlike mere lexical collections, this work endeavored to encompass both conceptual definitions and procedural knowledge, thereby bridging theoretical understanding with practical applications. Particularly notable is its innovative integration of textual explanations with graphic representations, especially concerning objects and manufacturing processes (Gürer, 2014). This dual modality aimed to enhance the user's grasp of complex processes by linking descriptive text with visual aids, a precursor to today's infographics. This methodological combination not only enriched the encyclopedic content but also anticipated modern multimodal communication strategies by recognizing the cognitive benefits of integrating textual and visual information.

2.3.2.2 Online encyclopedias and dictionaries

With the advent of the internet, encyclopedic knowledge transitioned from static printed volumes to dynamic, collaborative digital platforms. Wikipedia exemplifies this paradigm shift as the most widely used online encyclopedia, notable for its open editing model and multilingual accessibility (Wikipedia, 2022). Its strengths lie in providing a vast, interconnected network of hyperlinked articles that facilitate comprehensive topic exploration and cross-referencing. However, this model also introduces challenges, such as variable content quality, potential bias, and security vulnerabilities stemming from its collaborative nature.

Complementing Wikipedia's text-centric approach, there are emerging platforms that employ advanced visualization techniques to represent semantic relationships between concepts. For instance, Visuwords (<https://visuwords.com>) provides an interactive graphical interface depicting the lexical and conceptual connections between words using network models. As shown in Figure 2.5, the term "knowledge" is mapped onto a web of related words, illustrating how online textual

resources can leverage data visualization to enhance understanding and discovery. Such platforms exemplify the convergence of textual and visual knowledge representation, offering more intuitive and engaging ways to explore the semantic landscape of language.

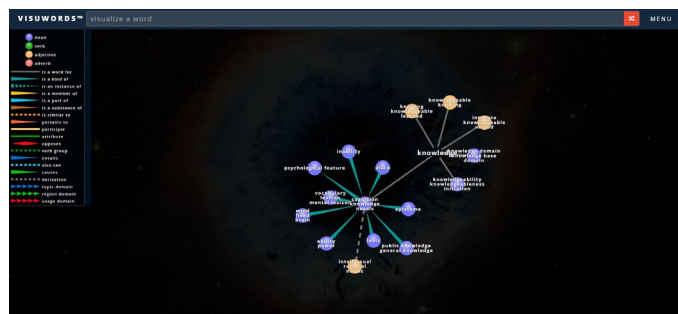


Figure 2.5 : Representation of the relations of the word "knowledge" with a network model on the Visuwords platform (Url-1).

As another example of a visual semantic tool, Visual Thesaurus stands out as an interactive lexical tool that maps the relationships between words through a dynamic, network-based interface (Figure 2.6). By organizing synonyms, antonyms, and related terms into a multidimensional spatial layout, the platform enhances the user's ability to perceive the semantic structure of language in a non-linear format. This approach not only facilitates vocabulary expansion but also contributes to the development of associative thinking by allowing users to observe how meanings are interconnected across grammatical categories.

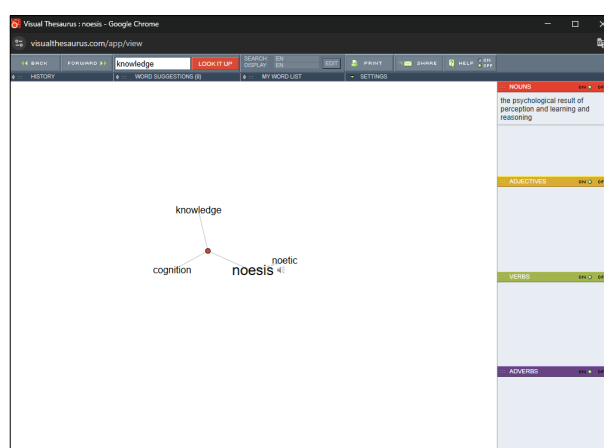


Figure 2.6 : Representation of the relations of the word "knowledge" with a network model on the Visual Thesaurus platform (Url-2).

Building upon similar principles, Lexipedia offers a visually structured and linguistically rich environment, further emphasizing the functional aspects of

language (Figure 2.7). While it shares Visual Thesaurus's network-based visualization, Lexipedia integrates syntactic roles and part-of-speech classifications more prominently, enabling users to explore not just the semantic but also the grammatical relationships between words. This nuanced integration of form and meaning supports deeper language comprehension and bridges conceptual and procedural dimensions of linguistic knowledge. The continuity between Visual Thesaurus and Lexipedia lies in their shared goal of transforming abstract lexical information into cognitively accessible, visual formats that foster pattern recognition and insight.

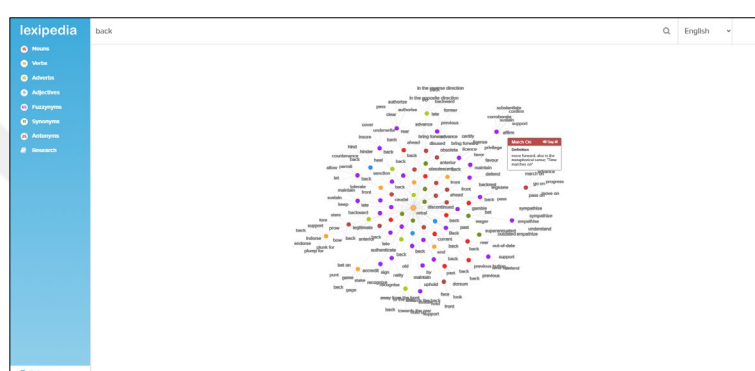


Figure 2.7 : Representation of the linguistic relationships with a network model on the Lexipedia platform (Url-3).

Extending the scope of visual knowledge representation beyond language, Kumu introduces a broader framework for mapping complex systems, ideas, and relationships across diverse domains (Figure 2.8). Unlike the more linguistically focused platforms, Kumu allows for the construction of interactive, multi-level knowledge networks that incorporate conceptual, procedural, and configurational data. Its flexibility in representing hierarchical structures, feedback loops, and relational dynamics makes it a valuable tool for analyzing systemic knowledge in architectural design, policy-making, or educational planning. In this way, Kumu complements the capabilities of Visual Thesaurus and Lexipedia by enabling users not only to visualize words and meanings, but also to navigate the complex interrelations that define knowledge ecosystems.

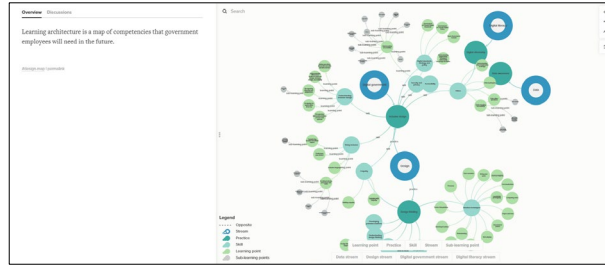


Figure 2.8 : Mapping the “design” discipline to build a learning architecture by Ksenia Cheinman (Url-4).

A further example is CiteSpace, a Java-based software developed for the visualization and analysis of trends and patterns within scientific literature, particularly in evolving research domains. It is commonly used to identify key developments, such as intellectual turning points or influential publications in a specific field. The application has been utilized in various case studies (Chen, 2006), offering a methodological framework for mapping scientific knowledge (Figure 2.9).

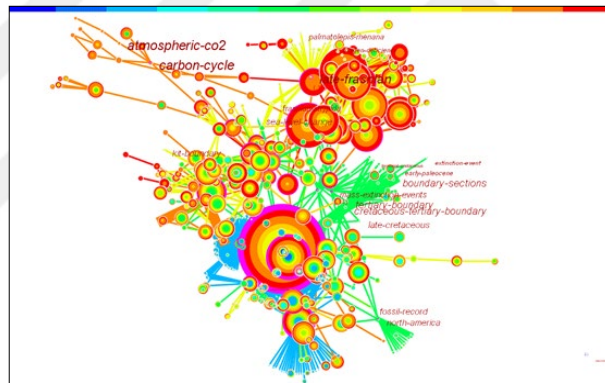


Figure 2.9 : Data visualization on CiteSpace platform (Url-5).

Another example of a network-based knowledge representation tool is Graph Commons, a collaborative platform designed for mapping and analyzing complex data relationships (Figure 2.10). Unlike word-focused lexical tools, Graph Commons enables users to construct interactive network maps that reveal systemic structures within social, organizational, or thematic datasets. Its utility lies in visualizing hidden connections, supporting critical data analysis, and aiding decision-making processes through a holistic understanding of networked information. By facilitating user-driven mapping practices, the platform contributes to knowledge production that is both participatory and structured around the logic of interconnectedness.

In conclusion, both visual and textual representations of knowledge play essential and complementary roles in the organization, communication, and comprehension of

The first example of the field of "humanistic computing," Index Thomisticus, was founded in 1949 by the Italian Father Roberto Busa and is regarded as the forerunner of digital humanities. A computerized index of words from the writings of the medieval philosopher Saint Thomas Aquinas is called Index Thomisticus. This 11 million word epic was published in 56 volumes during the 1970s. This effort, which opened the door to digital humanities, helped to spread the field of digital humanities study and set the precedent for initiatives such as the Perseus Digital Library (Adams and Gunn, 2012; Isomenger, 2018).

Özel and Öztemiz (2020) arranged the tools utilized in digital humanities into the following categories: digital publishing tools, data visualization tools, timeline tools, text editing and analysis tools, and 3D modeling tools. A-Frame, Kubity, Agisoft Metashape for 3D modeling tools; Drupal, Omeka, Plone, Scalar for digital publishing tools; Chronos Timeline, Historypin, QGIS, TimelineJS for timeline tools; Voyant Tools, MALLET for text editing and analysis tools; and Concordle, Wordle, Netlytic, Prism, Tableau, Gephi, Graphviz, D3 Data Driven Documents for data visualization tools are a few examples of these tools.

Furthermore included in this scope is VOSViewer, which has been employed extensively in recent scientific studies, which is used to create and visualize bibliometric networks. These networks can be built on the basis of citation, bibliographic coupling, co-citation, or co-authorship links, and they can involve journals, researchers or individual publications. Additionally, VOSViewer has text mining tools for building and visualizing networks of phrases that frequently appear together that are taken from scholarly publications (Figure 2.11).

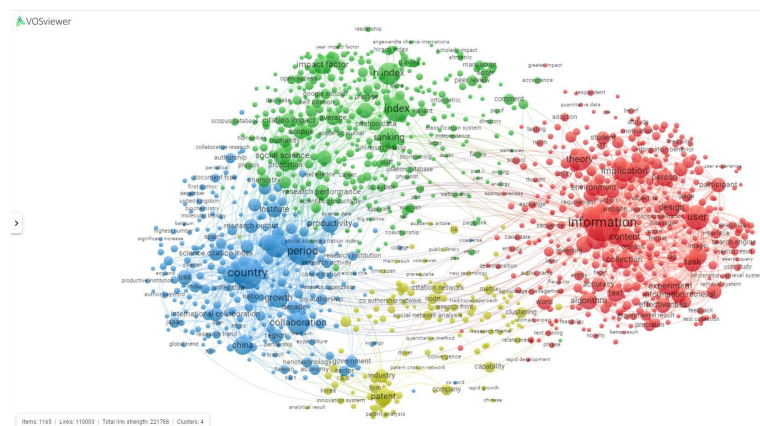


Figure 2.11 : A network representation obtained via VOSViewer software (Url-7).

2.5 Exploratory Learning via Interactive Data Visualization

In the digital age, the proliferation of data has transformed the landscape of education, necessitating innovative approaches to learning that go beyond traditional didactic methods. Traditional, teacher-centered, and didactic approaches often fall short in fostering deep understanding and the development of critical thinking skills, particularly in complex, multidimensional, and dynamic domains of knowledge (Biggs and Tang, 2011). In this context, there is an increasing need for pedagogical strategies that prioritize active learner engagement and promote self-directed exploration and meaning-making. Exploratory learning (closely related concepts: "discovery learning," "inquiry-based learning," "active learning," "constructivist learning") directly addresses this need by enabling learners to interact with knowledge through their own questions and curiosity, fostering understanding through direct engagement with content (Bruner, 1961).

Exploratory learning is rooted in constructivist theories, which posit that learners build new knowledge upon the foundation of prior understanding through active engagement (discovery) with their environment (Piaget, 2008). Piaget emphasized that children modify and reorganize their mental structures when engaged in learning environments that encourage open-ended interaction with knowledge. Although discovery-based methods are particularly beneficial in early childhood, research has shown that providing opportunities for active knowledge construction remains essential across all educational stages, especially when learners are supported in engaging meaningfully with the learning process (Aydınlı, 2016). In this paradigm, learning is not a passive absorption of information but an active process of hypothesis generation, testing and refinement.

Njoo (1994) argues that problem solving, discovery learning and inductive learning together form a comprehensive framework for defining exploratory learning. These approaches share a number of common characteristics, particularly in their emphasis on the constructive nature of learning. In each case, learners are encouraged to generate general principles or procedures on their own through active engagement. The underlying premise is that this active and constructive involvement facilitates the meaningful integration of new information into the learner's existing cognitive framework.

With the advancement of information and communication technologies, the digitalization of education has led to the emergence of a range of open-standard electronic tools and services, prompting a re-evaluation of the modes, contexts, and content of learning (Collins and Halverson, 2009). Notably, digital learning platforms such as MOOCs and distance education have considerably expanded learners' access to external resources and interactions. Within these technologically enriched virtual environments, the role of the learner is increasingly empowered, fostering greater autonomy and active engagement in the learning process. The overwhelming proliferation of learning resources and information has made it increasingly challenging for learners to identify content that is both relevant and valuable to their specific learning needs. This environment of information overload often hampers the efficiency of exploratory search. However, recent advances in artificial intelligence (particularly the emergence of large language models such as ChatGPT) have introduced new opportunities to enhance exploratory search strategies within educational contexts (Luo et al., 2025).

Thoughtfully designed external representations can facilitate learning and comprehension by making complex information more accessible and manageable. Interactive data visualizations serve as external cognitive artifacts that extend the learner's capacity to process complex information (Norman, 1993). Interactive data visualizations, as external representation tools, play a pivotal role in facilitating exploratory learning by enabling users to actively engage with complex information, uncover patterns through discovery and develop deeper critical thinking skills. By aligning with constructivist principles and leveraging technological advancements, educators can create dynamic learning environments that foster curiosity, adaptability and lifelong learning.

Interactive data visualizations offer a different way for users to engage with complex datasets. Instead of simply receiving information, users can explore data themselves, and notice patterns through visual interaction. This process encourages gradual discovery and helps users build a clearer understanding of the data. Compared to more fixed or structured methods, this approach supports a more flexible and user-driven learning experience. From a cognitive standpoint, interactive data visualizations play a crucial role in enhancing human cognitive abilities, particularly in pattern recognition and sense-making when dealing with large volumes of

information. According to Thomas and Cook (2005), visual analytics, an interdisciplinary field combining analytical reasoning with interactive visual interfaces, facilitates the sense-making process by integrating computational analysis with human cognitive capabilities. This integration enables users to process complex data more effectively, identify patterns, and derive meaningful insights. Furthermore, research by Sedig et al. (2014) emphasizes that interactive visualizations support complex cognitive activities by allowing users to engage in exploratory data analysis. These tools support the iterative process of exploration and discovery inherent in exploratory learning. They do so by enabling real-time manipulation of data representations such as filtering, linking, zooming, dynamic querying, refocusing, and reconfiguring. This direct engagement transforms static charts and graphs into dynamic analytical landscapes, significantly enhancing the potential for uncovering previously unobserved correlations or outliers. These interactive features transform data exploration into an active learning process, enhancing users' ability to make sense of complex information (Sedig et al., 2014). Ware (2013) explains that visualizations take advantage of the human visual system's natural ability to efficiently process graphical information, which helps reduce cognitive load compared to interpreting raw numerical data or large amounts of text. The interactive component further supports this by allowing users to externalize their thought processes, iteratively refining their understanding as they explore different facets of the data. As datasets continue to grow in complexity and volume, the role of such interactive and exploratory techniques will undoubtedly become increasingly central to effective scientific data interpretation and information acquisition.

In digital architecture, interactive data visualizations enable students to explore relationships and dynamic interactions among concepts. The proposed online encyclopedia supports exploratory learning by allowing users to investigate and manipulate data, uncover hidden and unexpected patterns and relationships, and develop a deeper understanding through non-linear, hands-on engagement with knowledge.

3. ARTIFICIAL INTELLIGENCE-ASSISTED TEXT MINING

As stated by Hearst (2003), text mining is the uncovering of previously undiscovered material by a computer by mechanically extracting information from various written resources. The linking of the extracted information together to produce new facts or new hypotheses to be examined further by more conventional modes of experimentation is a critical component. Text mining differs from what we are used to seeing in web search. The user is usually seeking for something that is already known and has been written by someone else when they search. The issue is ignoring all of the stuff that isn't immediately relevant to your needs in order to find the relevant information. The purpose of text mining is to discover previously unknown knowledge, something that no one knows and hence cannot be put down.

Hotho et al. (2005) express that Feldman & Dagan (1995) originally introduced text mining or knowledge discovery from text (KDT) as a term for the automated study of text. It links algorithms and techniques from knowledge discovery in databases (KDD), data mining, machine learning, and statistics with information retrieval, information extraction and natural language processing (NLP) techniques.

On the other hand, Gaikwad et al. (2014) expresses that according to the information retrieval basically there are four methods used:

- 1) Term Based Method (TBM)
- 2) Phrase Based Method (PBM)
- 3) Concept Based Method (CBM)
- 4) Pattern Taxonomy Method (PTM)

A document is evaluated using the term-based method, which has the advantages of effective computational performance and established theories for word weighting. Since phrases are less confusing and more discriminating than individual terms, phrase-based methods analyze documents on the basis of phrases. In concept-based terms are analyzed on sentence and document level. Documents are examined based on patterns when using the pattern taxonomy method.

Such analytical methods underpin many AI-assisted text mining applications, enabling effective processing and interpretation of textual data across multiple fields. Information retrieval systems utilize text mining techniques to improve search engine results and recommendation systems. Text classification and categorization techniques assist in tasks like spam filtering, sentiment analysis and topic classification. Sentiment analysis enables the identification and classification of sentiments expressed in text data, benefiting social media monitoring and brand reputation management. Text summarization and information extraction techniques aid in condensing large volumes of text and extracting key information for analysis.

Additionally, similar to web mining, text mining is frequently associated with data mining and hence seen as a particular type of data mining. The underlying database can be used to identify the main distinction between data and text mining. The examination of structured data is where traditional data mining got its start. According to relational database terminology, these are often in the first normal form. Because of this, each data field contains atomic values, or values that cannot be further subdivided. On the other hand, text documents are typically viewed as an unstructured base for analysis in the context of text mining. Grammatically speaking, a text is never entirely unstructured; rather, it exhibits two types of structure: implicit structure, which emerges from grammatical rules and linguistic conventions, and explicit structure, which is determined by content-related elements such as headings, subheadings, and paragraph organization (Hippner and Rentzmann, 2006).

This chapter is organized into three sequential sections, each of which contributes to a comprehensive framework for analyzing scholarly texts through computational methods. The first section, Text Preprocessing, outlines the essential preparatory steps required to transform the raw text from the scientific literature into a structured and analyzable dataset. This foundational phase ensures the quality and consistency of the data entering the analysis pipeline. The second section, titled Hierarchical Text Categorization and Clustering, details the central analytical stage of the methodology. In this part, advanced clustering algorithms are applied to the preprocessed text to uncover and map the hierarchical relationships between concepts and topics, revealing the underlying knowledge architecture of the field. Finally, the third section, Text Mining Applied to Scientific Studies, primarily presents a review of existing text mining research conducted on scientific

publications. It consolidates the methodological approach by illustrating how computational techniques have been utilized in prior studies to extract meaningful patterns, validate classification frameworks and derive insights that are often unattainable through traditional reading alone.

3.1 Text Preprocessing

The initial step of processing texts is text indexing. The process of turning a text into a list of words is known as text indexing. Textual data, in its raw, narrative form, is inherently unstructured and thus computationally challenging to process directly. Therefore, the foundational step in any text analysis is text indexing, a process that deconstructs sentence-based text into its constituent words or tokens, making it amenable to algorithmic manipulation (Jo, 2019).

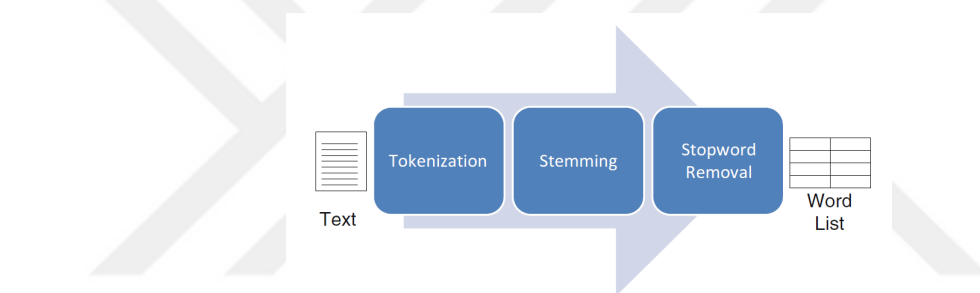


Figure 3.1 : The three steps of text indexing (Jo, 2019).

Figure 3.1 presents an illustration of the three fundamental phases in text indexing. Tokenization, the first phase, is the division of a text into tokens using white spaces or punctuation. The second phase is stemming, which involves employing grammatical rules to transform each token into its unique root form. The elimination of grammatical words like articles, conjunctions, and prepositions is the final phase, also known as stop-word removal. Tokenization is a prerequisite for the following two processes, but depending on the circumstances, stemming and stop-word removal may be substituted for one another.

As Witten (2004) highlights, automatic text classification has a wide range of practical applications, from document indexing and metadata extraction to word sense disambiguation and organizing vast web collections. The methodologies for achieving these applications, however, have evolved dramatically. Prior to the 1990s, the field relied primarily on "knowledge engineering," where classification rules

were manually derived from human experts and encoded into software. The subsequent, and now-dominant, paradigm is rooted in machine learning, which automates this process by enabling algorithms to infer classification patterns directly from a training set of pre-labeled documents, marking a decisive shift from expert-driven rules to data-driven models.

Text mining is a broad field encompassing a range of sub-disciplines, including information retrieval, document clustering, text summarization, and information extraction. For the purposes of this study, the research methodology leverages two specific text mining techniques; topic modeling and clustering algorithms, to analyze data extracted from the CUMINCAD database. The primary objective is to move beyond simple information retrieval and instead uncover the latent relationships between various concepts, ultimately classifying them into a meaningful hierarchical structure. The dataset extracted from CUMINCAD provides a rich foundation for this analysis, comprising a structured text file that includes article titles, keywords, abstracts, author information, publication years, and conference names for each entry.

3.2 Hierarchical Text Categorization and Clustering of Preprocessed Text Data

Following the foundational stage of text preprocessing, where the raw textual data was cleaned and structured, the methodology now advances to its central analytical phase. This section details the application of hierarchical text categorization and clustering techniques to the prepared dataset. The primary objective here is to move beyond the analysis of individual words or documents and instead uncover the latent, nested knowledge structures inherent within the entire corpus. By employing these methods, the research aims to automatically identify thematic groups and organize them into a meaningful, multi-level hierarchy that reflects the conceptual framework of the field.

The necessity for unsupervised learning approaches in text analysis arises from fundamental limitations inherent in supervised classification methods. Supervised techniques presuppose the existence of an extensive, pre-labeled training corpus, a resource that is often impractical to create due to the immense labor required for manual annotation. This problem becomes acute in two key scenarios. First, in nascent fields of study, a stable classification schema may not exist, making the task

of pre-labeling conceptually impossible. Second, even in established domains, if the subject matter is subjective or open to varied interpretations, expert labeling can produce inconsistent and unreliable training data. Therefore, unsupervised methods like clustering offer a vital alternative, enabling the automatic discovery of latent structures within the data without the prerequisite of human-annotated ground truth (Shafiabady et al., 2016).

In Figure 3.2, a decision tree has been created to demonstrate how answering a few simple questions can guide towards an appropriate text mining solution. This decision tree branches out and its leaf nodes lead to main branches. A text mining solution typically involves multiple transitions between different levels of data processing, starting from raw input documents and progressing towards fully digitized text (Miner et al., 2012). Within the scope of this study, answering the questions in Figure 3.2 leads to the main branch of “Clustering”.

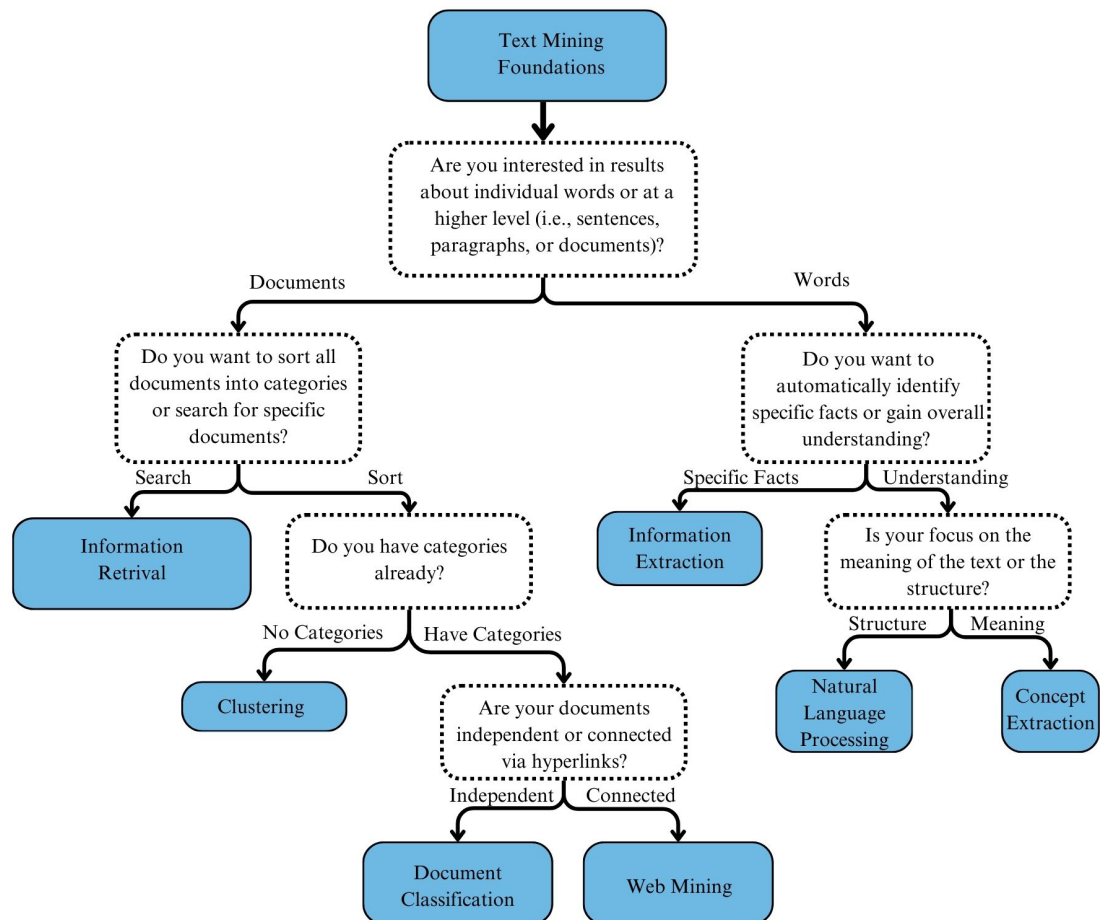


Figure 3.2 : A decision tree for identifying application areas in accordance with text mining resources (Miner et al., 2012).

Unsupervised learning techniques, particularly K-means clustering, are instrumental in uncovering the latent structures and thematic patterns within the extensive corpus of digital architecture articles. By applying the K-means algorithm to the vector representations of article titles and keywords, this study can systematically group documents based on their semantic similarity. This process of automatic categorization is especially valuable as it operates without the need for pre-defined labels or manual intervention. Consequently, it allows for the data-driven discovery of conceptual clusters and relationships that may not be immediately apparent through traditional qualitative analysis, thereby enabling a more objective and scalable approach to organizing a large body of scholarly work.

Manual organization and labeling of a significant number of articles can be labor-intensive and intellectually challenging. This procedure is streamlined by unsupervised learning, especially K-means clustering, which automatically groups articles into clusters, saving a significant amount of time and effort. K-means is among the most widely used clustering algorithms in data mining and statistics, as noted by Duda and Hart (1973) and Hartigan (1975). Originally developed within the field of statistics, this method is known for its straightforward implementation and suitability for large datasets. In particular, K-means has proven to be highly effective in text clustering applications (Hotho et al., 2005).

The application of unsupervised learning algorithms, and specifically K-means clustering, is particularly well-suited to the dynamic and often nascent field of digital architecture for several critical reasons. First, in emerging research areas where established classification schemas may be lacking, K-means operates without the need for predefined categories. This allows for the organic, data-driven discovery of thematic clusters in uncharted intellectual territories. Second, the algorithm effectively accommodates the high degree of variability and subjectivity inherent in digital architecture discourse, where diverse perspectives and interpretations are common. By grouping documents based on semantic proximity rather than rigid rules, it ensures robust clustering even when faced with heterogeneous or opinion-based textual material.

Furthermore, the scalability of K-means offers a significant practical advantage. Given the large and ever-growing volume of publications in digital architecture, this computational efficiency provides a feasible solution for large-scale data analysis. By

applying this method to the titles and keywords of scholarly articles, researchers can achieve a comprehensive, macro-level understanding of the field's research landscape. Ultimately, this approach not only identifies dominant research trends and latent connections but also provides a structured foundation for informing future investigations.

In addition to K-means clustering algorithms, the TF-IDF method (Salton & Buckley, 1988), which measures the importance and frequency of a term within a document, is also utilized in this study. In the TF-IDF method, TF (Term Frequency) indicates the frequency of a specific word in the document, while IDF (Inverse Document Frequency) is applied to determine the frequency of a word across multiple documents, thereby identifying whether the word is a term or merely a common word such as a conjunction.

The TF (Term Frequency) formula (3.1, 3.2, 3.3) (Salton and Buckley, 1988):

$$TF_{i,j} = \frac{n_{ij}}{\sum_k n_{k,j}}, \quad (3.1)$$

The IDF (Inverse Document Frequency) formula:

$$IDF_i = \log \frac{n_d}{\{d: t_i \in d\}} \quad (3.2)$$

The TF-IDF (Term Frequency-Inverse Document Frequency) formula:

$$(TF - IDF)_{i,j} = TF_{i,j} \times IDF_i \quad (3.3)$$

By applying a logarithmic scale to the ratio of the total number of documents to the number of documents containing the term, IDF assigns a higher weight to terms that are more unique and thus have greater discriminatory power.

By multiplying these two values, the final TF-IDF score assigns the highest weight to terms that are frequent in a specific document (high TF) but rare across the collection as a whole (high IDF). This makes it an exceptionally effective method for identifying keywords that uniquely characterize a document's content, which is crucial for tasks such as information retrieval and text clustering. This method is used to examine whether a word is associated with a specific topic or area of

expertise and, if so, to analyze the characteristic features of that field. It evaluates the frequency of specific terms to determine their degree of association with the field of expertise.

Both K-means clustering algorithms and the TF-IDF method can be employed using the Python programming language. The scikit-learn library in Python provides the k-Means and TfidfVectorizer classes for this purpose.

The structured algorithm description below provides a step-by-step guide for creating the clusters for the encyclopedia using CUMINCAD database:

Algorithm: K-Means Clustering with TF-IDF Vectorization.

Input: A corpus of preprocessed text documents.

Output: A set of document clusters with their characteristic term distributions.

Data Preprocessing: The raw text corpus is subjected to a standard natural language processing pipeline, including tokenization, lemmatization, and the removal of stop words, to produce a clean set of meaningful tokens.

TF-IDF Vectorization: The preprocessed corpus is transformed into a numerical document-term matrix using the TF-IDF weighting scheme. Each document is thereby represented as a high-dimensional vector, where each dimension corresponds to a term and the value is its TF-IDF score.

Centroid Initialization: A pre-determined number of clusters, K (in this study, approximately 30), is specified. The initial K cluster centroids are randomly positioned within the vector space.

Iterative Assignment and Update: The algorithm enters an iterative loop that continues until convergence is achieved (i.e., when cluster assignments no longer change significantly between iterations):

- a. Assignment Step: Each document vector is assigned to the cluster whose centroid is nearest.
- b. Update Step: The position of each cluster centroid is recalculated as the mean (i.e., the geometric center) of all document vectors currently assigned to that cluster.

Output and Analysis: Upon convergence, the algorithm outputs the final cluster assignments for each document. The term distributions for each cluster are then analyzed to interpret their dominant themes.

3.3 Text Mining Applied to Scientific Studies

In recent years, text mining has gained prominence as a powerful approach for analyzing the rapidly expanding volume of unstructured textual data in scientific research, reflecting the exponential increase in scientific output originally identified by Price (1965) and subsequently substantiated through bibliometric analyses (Tabah, 1999). The ongoing digitization of human knowledge across diverse formats such as articles, blogs, multimedia, and social media has made it increasingly difficult to locate relevant and meaningful information. Conventional tools like keyword-based search and hyperlink navigation offer basic access to digital content but often fail to reveal the deeper conceptual structures that connect documents. An alternative approach involves organizing and accessing information based on thematic structures, allowing users to focus on specific topics, observe their progression over time, and identify meaningful connections. This method enhances both comprehension and discovery by prioritizing conceptual understanding over surface-level retrieval (Blei, 2012).

In the context of scientific literature, the challenge of navigating the ever-expanding body of research has made text mining an essential methodological tool. By applying computational techniques to large corpora of academic texts, researchers can extract patterns, detect trends, and uncover latent structures that are not easily visible through traditional reading or manual review.

Text mining, also known as knowledge discovery from text (KDT), was first introduced by Feldman and Dagan (1995) and refers to the computer-assisted analysis of textual data. It integrates techniques from information retrieval, information extraction and natural language processing (NLP) with methods drawn from knowledge discovery in databases (KDD), data mining, machine learning and statistics to extract patterns, structures and insights from large volumes of unstructured text. In scientific contexts, it enables researchers to extract latent structures from corpora composed of abstracts, full texts, metadata and citations. Its applications include topic modeling, document clustering, keyword extraction and sentiment analysis, each of which contributes to a deeper understanding of the thematic landscape of a research field. Text mining supports tasks such as identifying frequently occurring terms, discovering co-authorship or citation networks, mapping

the evolution of specific research topics and clustering documents based on semantic similarity. Various text mining techniques are employed to uncover patterns within large text corpora. Miner et al. (2012) outline seven core practice areas that emerge from addressing key resource and objective questions faced by text mining practitioners. These areas reflect the intersection of text mining with related disciplines and encompass a wide range of analytical approaches (Figure 3.3). They include information retrieval, which focuses on storing and accessing text through tools like search engines; document clustering and classification, which involve grouping or categorizing text elements using unsupervised and supervised learning methods, respectively; and web mining, which emphasizes the analysis of large-scale, interconnected online data. Additionally, information extraction transforms unstructured text into structured formats by identifying entities and relationships, while natural language processing supports foundational linguistic analysis such as part-of-speech tagging. Finally, concept extraction deals with identifying and grouping semantically similar terms, contributing to a deeper understanding of textual content across domains. Each of these approaches serves a distinct purpose, from tracing conceptual developments within a field to identifying emerging interdisciplinary connections, thereby contributing to a more systematic and scalable understanding of scientific knowledge production.

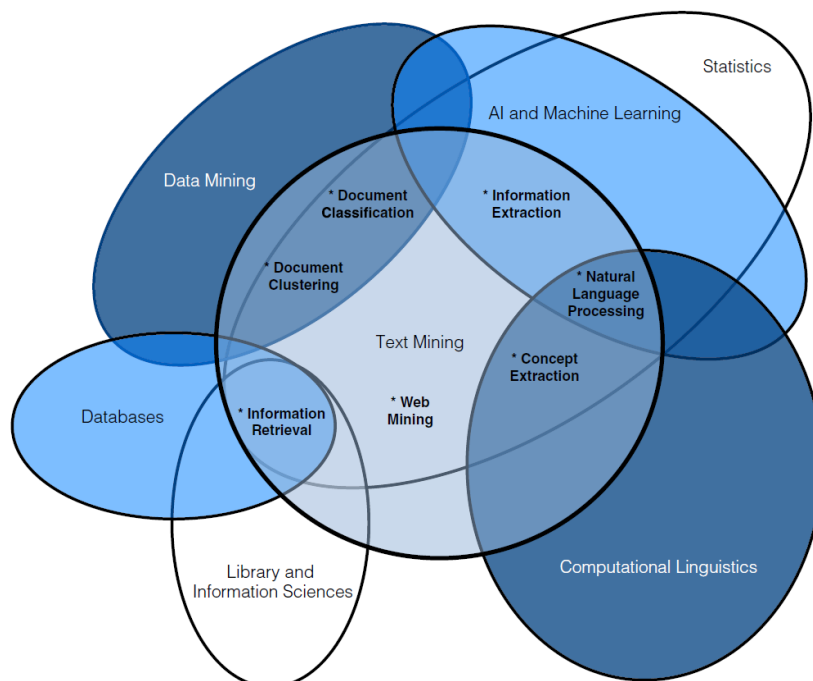


Figure 3.3 : A Venn diagram illustrating the overlap between text mining and six associated disciplines (Miner et al, 2012).

One of the most prevalent text mining techniques in scientific knowledge discovery is topic modeling, particularly Latent Dirichlet Allocation (LDA). Topic modeling algorithms uncover abstract topics within a corpus by identifying word co-occurrence patterns, offering a probabilistic framework for understanding the thematic structure of a document collection (Blei et al, 2003). The underlying concept of Latent Dirichlet Allocation (LDA) is that a set of topics, each defined as a probabilistic distribution over words, is present throughout the entire document collection (as illustrated on the far left). The generative process for each document involves first selecting a topic distribution (as shown in the histogram on the right). Then, for every word in the document, a specific topic is sampled (represented by the colored coins), and a word is subsequently drawn from the word distribution associated with that topic (Figure 3.4).

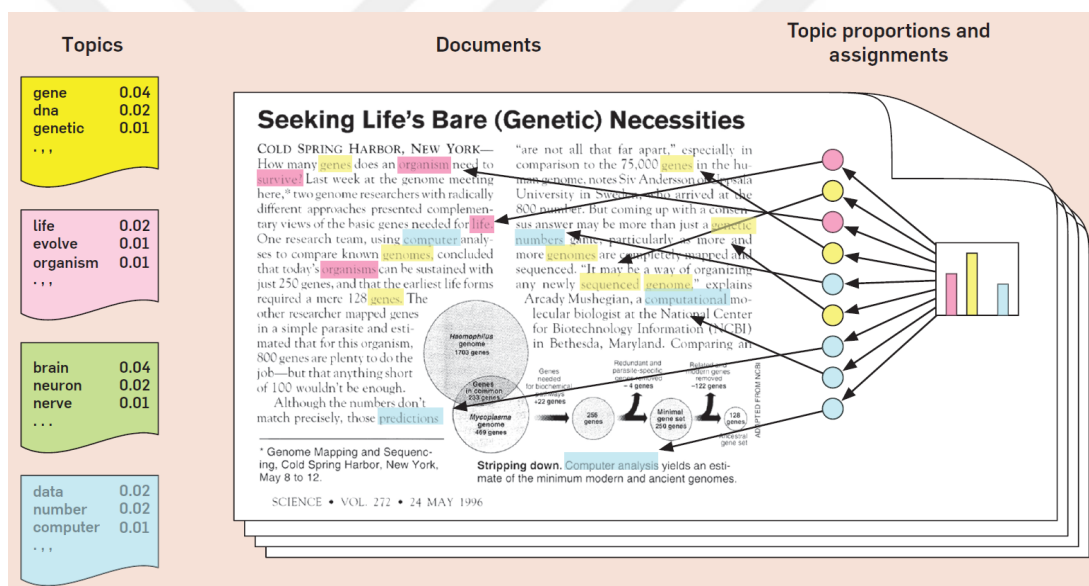


Figure 3.4 : The underlying concept of Latent Dirichlet Allocation (LDA) (Blei, 2012).

Another essential technique in the text mining of scientific literature is cluster analysis, which identifies latent structures and thematic patterns by grouping data points, such as documents, based on their inherent similarities, without relying on predefined labels. In the context of text mining applied to scientific studies, cluster analysis operates without predefined category labels, that is, it does not rely on prior class identifiers. This lack of labeled data distinguishes clustering (an unsupervised learning method) from classification or discriminant analysis, which are forms of supervised learning. The primary goal of clustering is to uncover inherent structure

within the data, making it an exploratory technique by nature. Clustering has a long-standing and diverse history across many scientific disciplines. One of the earliest and most widely used algorithms, K-means, was introduced in 1955. Despite the development of thousands of clustering algorithms since then, K-means remains prevalent, highlighting both the challenges in creating a universally applicable clustering algorithm and the inherently ill-defined nature of clustering itself (Jain, 2010). Cluster analysis is frequently coupled with term-weighting schemes such as Term Frequency–Inverse Document Frequency (TF-IDF), which transforms textual data into numerical feature vectors based on the relevance of terms across documents. These vectorized representations enable algorithms like K-means to measure semantic similarity and group documents accordingly. The use of TF-IDF not only enhances the discriminative power of clustering but also facilitates the identification of coherent thematic groupings within large-scale scholarly corpora (Aggarwal and Zhai, 2012).

The text preprocessing stage, comprising tokenization, stemming, lemmatization, stop-word removal, is critical to ensure data consistency and model accuracy. Texts from scientific articles often contain domain-specific language, necessitating careful corpus preparation to maintain the integrity of disciplinary semantics.

To effectively carry out these critical preprocessing and analysis tasks, researchers can leverage a powerful ecosystem of open-source libraries and dedicated software tools. Several open-source libraries and tools facilitate text mining for scientific research. Python libraries such as NLTK, spaCy, and Gensim are widely used for building pipelines, while Scikit-learn provides scalable machine learning functionality. For more specialized needs, platforms like VOSviewer, CiteSpace, and KNIME offer GUI-based environments tailored for bibliometric analysis. These tools support visual analytics that complement text mining outputs, enhancing interpretability and user engagement in scientific studies.

Text mining is particularly powerful when integrated with scientific data visualization. Graphical visualization often conveys information more clearly, comprehensively and quickly than purely text-based descriptions, making it a valuable tool for exploring and analyzing large document collections (Hotho et al, 2005). For example, topic modeling results can be represented as interactive sunburst charts or circle packing diagrams using libraries like D3.js, allowing users to explore

thematic hierarchies intuitively. Hierarchical edge bundling or radial trees can represent relationships among subdomains, enabling a dynamic and exploratory form of knowledge interaction. These representations are not only informative but also pedagogically valuable, offering learners an entry point into complex research landscapes.

Taken collectively, these studies reveal that text mining constitutes a vital methodological approach for mapping, analyzing, and interpreting scientific literature. Its combination with advanced visualization techniques supports exploratory learning and knowledge construction, particularly in interdisciplinary fields where traditional classification systems fall short. When applied to curated datasets like CUMINCAD, text mining not only supports research synthesis but also contributes to the design of interactive knowledge environments, such as online encyclopedias or learning platforms. Moreover, by enabling the extraction and organization of conceptual, procedural, visual, and configurational knowledge, text mining fosters exploratory learning processes that are essential for navigating and constructing meaning within complex and dynamic knowledge domains.



4. DATA VISUALIZATION AS A LEARNING TOOL: MAPPING AND CATEGORIZATION OF DIGITAL ARCHITECTURE KNOWLEDGE

4.1 Knowledge Classification in Digital Architecture

Digital architecture is inherently interdisciplinary, drawing upon principles from computer science, engineering, art, design, and the humanities. Consequently, the knowledge embedded in digital architecture practices is multifaceted. It encompasses theoretical frameworks, computational methods, digital tools, fabrication techniques and conceptual approaches to spatial organization and representation. Given this complexity, effective classification must capture the diverse dimensions of knowledge production, application and interpretation within the evolving landscape of digital architecture discourse.

As the volume of digital architecture knowledge expands, researchers and practitioners are increasingly confronted with the challenge of navigating complex and often fragmented information landscapes. A systematic classification not only aids in knowledge retrieval and discovery but also enhances communication between different stakeholders such as academics, students, and professionals.

The literature presents multiple categorization approaches to map the multifaceted components of digital architecture knowledge. Kolarevic (2000) offers a foundational categorization of digital architectures by linking them directly to their computational underpinnings. He distinguishes between various types based on the core principles or algorithms that drive their form generation and transformation processes. These include:

1. Topological architectures, which stem from concepts of continuous deformation in topological space;
2. Isomorphic architectures, based on the transformation of surfaces that retain a structural likeness;
3. Animate architectures, driven by motion kinematics and dynamic systems;

4. Metamorphic architectures, utilizing keyshape animation techniques to morph between forms;
5. Parametric architectures, defined by variable-driven design logics;
6. Evolutionary architectures, which incorporate genetic algorithms and evolutionary computation to generate optimized solutions.

Kolarevic's classification is not just a typology of aesthetic styles or design philosophies; rather, it is a computational taxonomy. It reflects how different digital design approaches emerge from specific algorithmic principles.

On the other hand, in the book *Atlas of Digital Architecture*, edited by Hovestadt et al. (2020), the expansive domain of digital architecture is systematically categorized into six overarching thematic sections, each reflecting distinct epistemological and methodological dimensions of the field:

1. **The Design – Creating the Geometries of Architectural Artefacts:** This section encompasses fundamental methods and tools for geometric and spatial generation in architecture, including 3D Modelling, Digital Data Acquisition, Digital Design Strategies, Computer-Aided Design (CAD), Generative Methods and the theoretical underpinnings of Graphs & Graphics.
2. **The Image – Visualising Architecture:** Focused on visual representation, this category includes techniques and technologies for architectural visualization, such as Image & Colour theory, Rendering techniques, and strategies for Visualisation that enhance perceptual comprehension of design intent.
3. **Language – The Abstraction of Architecture:** This part explores the textual and coded forms of architectural expression, emphasizing the role of Text, Typography & Layout, Scripting languages, and the integration of Writing & Code as both design media and communicative tools.
4. **Matter & Logic – The Physical Representation of Architecture:** Addressing the translation from digital to physical, this classification includes Digital Manufacturing techniques, Model Making, and 3D Printing, thereby linking computational design with material execution.
5. **Logistics – The Dynamic Representation of Architecture:** This section is concerned with the temporal, performative, and systemic aspects of architecture,

incorporating tools and frameworks such as Virtual and Augmented Reality, Simulation, Geographic Information Systems (GIS), Building Information Modelling (BIM), Digital Cities and the application of Big Data & Machine Learning.

6. Coexistence – Interfaces and Modes of Collaboration Between Information Technology and Architecture: Lastly, this category explores the socio-technical intersections and collaborative dynamics between architecture and information technology. Topics include branding in digital contexts (Being a 'Brand'), the integration of the Internet of Things (IoT), collaborative platforms and workflows (Collaboration), and concerns around Privacy & Security.

A comprehensive review published in 2020 by Caetano and Leitão systematically examines the evolution of computational design (CD) technologies within architectural theory and practice through a four-stage methodology:

1. Historical Literature Mapping: The first stage entailed an extensive review of the architectural literature on CD, spanning from the 1960s to the present. Key contributions that significantly influenced the integration of computational approaches in architecture were identified, alongside an analysis of leading conferences and journals responsible for shaping the discourse.
2. Technological Contextualization: The second phase focused on situating these developments within broader technological advancements, highlighting digital tools that were particularly suited to architectural processes and the technical innovations they introduced.
3. Synthesis of Theory, Technology, and Practice: In the third stage, the review established parallels between theoretical advancements, technological availability and real-world applications of CD techniques. This alignment emphasized the reciprocal relationship between conceptual evolution and practical implementation.
4. Temporal Structuring and Critical Reflection: The final phase involved constructing dual timelines, one tracing technological and methodological milestones, the other organizing significant theoretical contributions. This was accompanied by a critical reflection on CD's influence on architectural thinking

and practice, addressing both its transformative potentials and its inherent limitations.

These perspectives reflect the increasing theoretical complexity and epistemological depth of the field. Together, they contribute to a more rigorous classification of digital architecture knowledge, extending beyond tools to encompass cognitive processes, design strategies, and material practices.

From an educational perspective, Soliman et al. (2019) examined how computer technologies are integrated into architectural education by systematically analyzing the curricula of 20 architecture schools. Rather than focusing on specific software tools, which may vary across institutions and change over time, the authors proposed a classification based on the functional role of digital applications within the curriculum. They determined the comprehensive set of categories covered in the academic programs of the 20 schools:

1. 2D & 3D Representation;
2. Building Information Modeling (BIM);
3. Parametric Design;
4. Geographic Information Systems (GIS);
5. Digital Fabrication;
6. Simulation;
7. Environmental Technology & Building Technology;
8. Programming, Coding, and Scripting;
9. Communication.

Based on the curricular analysis conducted by Soliman et al. (2019), this functional categorization provides a valuable contribution to the ongoing discussion on how digital architectural knowledge can be systematically organized. Rather than organizing knowledge around transient software platforms, it emphasizes the functional dimensions of digital competencies, such as representation, simulation or fabrication, which are more enduring and conceptually grounded. In doing so, this classification supports the development of a systematic body of knowledge in digital architecture, helping to distinguish between various domains and their

epistemological roles within architectural education and practice. It reflects a shift from tool-centric training to an approach that prioritizes the underlying cognitive and procedural knowledge needed to engage critically with digital technologies. As such, this model aligns closely with efforts to construct hierarchical, conceptual taxonomies that capture the multifaceted nature of digital architecture knowledge.

Furthermore, the “Research Areas” outlined on the official website of the Architectural Design Computing graduate program at Istanbul Technical University (ITU) present a comprehensive framework for articulating the multifaceted nature of digital architecture (ITU MBL, 2025). The classification encompasses a wide array of research areas situated at the intersection of architecture and information technologies. These include computational design methods, artificial intelligence applications, robotic tectonics, automation and interactive systems; generative approaches such as shape grammars, fractals and evolutionary strategies; and data-driven methods like data mining and spatial data processing. Additionally, it incorporates emerging paradigms shaped by advancements in digital technologies such as cyberspace, virtual and augmented reality (VR/AR), mixed reality (MR), extended reality (XR), interface and interaction design, embedded environments, and the metaverse as well as studies on multimedia representation, remote sensing, cognitive processes and building information modeling (BIM). Crucially, this structure shifts the focus from software-centric instruction to a more conceptual and functional organization of knowledge. It emphasizes methodologies, cognitive processes, representational systems, and the technological infrastructure that support digital architectural practice. In doing so, it reflects a deeper epistemological engagement with how architectural knowledge is produced, mediated and transformed through digital means. It captures both the intellectual diversity of the field and its operational complexity, making it a valuable reference for curricular development, interdisciplinary research and knowledge management in architectural education and practice.

All of these categorizations have been developed manually rather than through text mining approaches. Existing studies that focus on digital architecture, architectural computing or computational design and employ text mining techniques remain relatively limited and predominantly utilize earlier methods. This thesis distinguishes itself in this regard by leveraging the widespread application of artificial intelligence

to perform text mining more effectively, enabling the extraction of meaningful and hierarchical classifications from vast bodies of information.

Additionally, data mining studies in the field of digital architecture conducted using earlier methods are as follows:

Using the CUMINCAD dataset, Turk et al. (2001) employed text analysis techniques to automatically classify CAAD papers. Although interpreting the resulting categories still required human reasoning, the outcomes were ultimately unsatisfactory, as the generated classifications significantly differed from those created by human experts during conference sessions (Martens et al, 2003).

Chiu et al. (2002) identified thirteen thematic clusters within the research: Design Methods and Models, Design Cognition, Collaborative Design, Information Systems, Design Media and Human-Computer Interaction (HCI), Precedents and Prototypes, Shape Studies and Knowledge Representation, Generative Systems, Virtual Reality (VR) and Virtual Environments (VE), Simulation, Prediction and Evaluation, Regional Information and Design Education. In this research, the data mining process was structured chronologically and thematically. First, the dataset was divided by publication years. Then, domain experts selected representative sample papers tied to specific topics. Abstracts were scanned to extract keywords, which were used to construct semantic relationships. These relationships enabled broader dataset analysis to classify or link papers by topic. The process concluded with an evaluation and refinement of the semantic framework. For instance, collaborative design papers from 2001 were analyzed to trace topic development over time, with keyword selection guided by frequency within abstracts (Chiu et al, 2002).

Martens et al. (2003) explored the conditions for developing a keyword-based classification system for CAAD publications. Drawing on the frequency of keyword entries in the CUMINCAD database, they identified approximately 30 core thematic categories, defined 5–10 subtopics for each, and selected representative papers accordingly. The initial list of categories was derived from the frequency of keyword entries in the CUMINCAD repository and included: 2D Representation, 3D City Modeling, 3D Modeling, Animation, Artificial Intelligence, Case Based Reasoning, Collaborative Design, Communication, Computer Integrated Construction, Constraint Based Design, Database Systems, Design Methodology, Design Process,

Digital Design Education, Digital Media, Environmental Simulation, Generative Design, Human-Computer Interaction, Image Processing, Interactive Design, Knowledge Modeling (KM), Learning Environment, Object Oriented Modeling, Performance Simulation, Shape Grammars, Virtual Design Studio, Virtual Environments, Virtual Reality, Visualization, Web Design. This effort stands as one of the early data-driven attempts to structure and enhance accessibility to knowledge in the domain of digital architecture.

All of these categorizations have been developed manually rather than through text mining approaches. Existing studies that focus on digital architecture, architectural computing or computational design and employ text mining techniques remain relatively limited and predominantly utilize earlier methods. This is where the evolution of knowledge discovery tools becomes relevant. The historical trajectory of digital architecture discourse, as chronicled in seminal conference series like CAAD, ACADIA, and eCAADe, reveals a progression from basic computational aids to complex generative and analytical systems. This evolution is well-documented in foundational studies. Early research by Martens and Turk (2002) established the CUMINCAD database in response to the challenge of dispersed and unindexed digital design literature. Their work emphasized the importance of structured repositories and hierarchical classification systems, which are closely related to the core issues explored in this thesis. In 2015, Celani and Franco provided a historical account of the development of CAAD conferences and affiliated organizations. Their analysis illustrated how the field gradually shifted from theoretical explorations toward more pragmatic and tool-oriented approaches. More recently, in 2019, Güzelci et al. conducted a content analysis of a large body of publications. Their findings identified patterns of emerging, fading, and stabilizing concepts within computational design, highlighting how text mining can reveal the evolving nature of the discipline.

This evolution has culminated in the emergence of contemporary AI-driven platforms such as Scholar GPT, Consensus and Scopus AI, which offer powerful capabilities for synthesizing research and answering specific queries through natural language processing. These tools excel at providing summarized, text-based outputs for targeted information retrieval. However, their primary function is one of "answering," which differs fundamentally from the objective of this study. The

proposed online encyclopedia is not designed to provide definitive answers but to foster exploratory learning and structural understanding based on a novel classification.

While AI-driven search engines provide linear paths to information, the framework developed here visualizes the entire knowledge domain as a navigable and relational map. By classifying knowledge into four distinct types (conceptual, procedural, visual, and configurational) it encourages users to understand the "how" and "why" behind concepts, not just the "what." This approach, which prioritizes the visualization of hierarchical and interconnected structures, creates a pedagogical environment that supports non-linear exploration and serendipitous discovery, a feature absent in query-based AI tools. Thus, this research positions itself not as an alternative to AI search engines, but as a complementary pedagogical tool that addresses the unique cognitive demands of learning within a complex, design-oriented field like digital architecture.

The exploration of knowledge classification within digital architecture reveals a consistent scholarly effort to structure its multifaceted and rapidly evolving domain, primarily through manual, expert-driven categorization schemes. These frameworks, while valuable for providing conceptual clarity and pedagogical structure, highlight a significant gap: the limited application and efficacy of earlier text mining techniques in automatically deriving comprehensive and nuanced classifications from the vast body of digital architecture literature. In response to these limitations, this thesis introduces a novel methodology that utilizes contemporary, AI-driven text mining. The aim is to construct a more meaningful, hierarchical, scalable, and simultaneously visualized classification framework that reveals the relationships between concepts, designed to systematically organize and improve accessibility to the ever-expanding knowledge landscape of digital architecture.

4.2 Data Visualization as a Learning Tool

Data visualization refers to the graphical representation of data, enabling the communication of complex information in a concise and accessible manner that would otherwise be challenging to interpret through raw numerical values alone. One of its primary advantages lies in its ability to significantly reduce the time required for both data presentation and interpretation. While the data may require

preprocessing before visualization, the resulting graphical formats enhance clarity and comprehension. Moreover, data visualization aids in identifying patterns and predicting trends, which can be instrumental in informed decision-making. It also has the potential to reveal insights or anomalies that might be overlooked in the absence of visual tools. For these reasons, data visualization is extensively used by scientists, engineers and analysts who rely on it for reporting, exploratory analysis and effective communication. Its capacity to streamline analysis processes not only increases efficiency but also contributes to cost savings by reducing the time and effort needed for comprehensive data evaluation (Pathak and Pathak, 2020).

This study employs data visualization methods to allow to explore information interactively. Rather than acquiring digital architecture knowledge through traditional didactic methods, users are encouraged to discover concepts through various hierarchical diagrams and interactive visuals. This approach fosters direct exploration, meaning-making and information acquisition.

The use of text mining techniques, such as topic modeling and clustering algorithms, provides users with the opportunity to engage in data analysis and establish connections between concepts in digital architecture. This active research process not only enables users to acquire knowledge but also to structure and deeply comprehend it. Such activities contribute to the development of problem-solving and analytical thinking skills, which are essential components of exploratory learning methodology.

In this study, users are encouraged to explore datasets through different visual representation methods. These visualizations assist users in structuring their learned material visually and improving their understanding of the relationships between concepts. As users explore digital architecture knowledge, they are also afforded the chance to discover new connections and make mistakes through the use of diagrams and relationships. This iterative process allows users to have a deeper and more meaningful learning experience as they progress along their educational journey.

Users can explore the knowledge at their own pace through different visual representations and datasets. This individualized approach ensures that each user experiences a unique learning process by uncovering different relationships. Exploratory learning, in this context, encourages users to engage with content that aligns with their own interests, thus personalizing the learning experience.

The study aims to visualize the relationships between concepts within digital architecture, allowing users to explore these connections rather than focusing on isolated pieces of information. This broadens the scope of exploration, facilitating a more comprehensive and effective learning process that is characteristic of exploratory learning.

Utilizing tools like mind mapping software, concept mapping platforms and diagramming applications, users are guided to explore and categorize knowledge in ways that mirror real-world design practices. Additionally, the integration of conceptual, procedural, visual, and configurational knowledge in teaching frameworks allows students to grasp the interconnected nature of design and knowledge systems.

Diverse visualization methods have been employed through the use of the D3.js platform to facilitate the dynamic and interactive representation of complex architectural data structures (Figure 4.1). The D3.js (Data-Driven Documents) library was selected as the primary platform for data visualization in this research due to its robust capacity to generate dynamic, interactive and highly customizable visual representations of complex datasets. One of the core advantages of D3.js lies in its ability to bind arbitrary data to the Document Object Model (DOM) and apply data-driven transformations to the document. This feature enables the creation of responsive and richly interactive graphics that allow users to engage directly with the content, facilitating exploratory analysis and deeper cognitive processing. Compared to other visualization tools, D3.js offers a high degree of flexibility in both design and function, supporting a wide range of formats including hierarchical trees, force-directed graphs, radial layouts and zoomable structures suited for the representation of layered and interconnected digital architecture knowledge. Furthermore, D3.js is highly compatible with various standard data formats, such as JSON (JavaScript Object Notation), CSV (Comma-Separated Values) and TSV (Tab-Separated Values). This versatility allows for seamless integration of external datasets, including those derived from digital knowledge repositories, bibliographic databases and semantic classification models. The ability to parse, manipulate and visualize these data types directly within the browser environment supports an efficient and streamlined workflow for both development and iterative refinement. Additionally, the open-source nature and active development community of D3.js ensure

continuous support and adaptation, while its integration with standard web technologies (HTML, SVG, CSS, and JavaScript) makes it relatively accessible and efficient for rapid prototyping (Cook, 2022). These qualities align closely with the goals of this thesis, which aims to explore, classify and visually communicate digital architecture knowledge in an engaging, flexible, and systematic manner. Such visualizations can support learning by providing an external representation similar to the concept maps that users construct in their minds. This 'whole-to-part' or 'part-to-whole' exploratory process enables users to better understand the connections between concepts and the role of each element within the overall system. The interactive nature of these D3.js-generated graphics inhibits passive information reception by users. Through interactions such as hovering, clicking or dragging, users can dynamically manipulate the visualizations, explore alternative scenarios, and engage with the evolving nature of conceptual structures in a more active and exploratory manner.

These visualizations are not only instrumental in revealing latent patterns and relationships within the dataset but also serve as cognitive tools that support the analytical and interpretive processes of the user. The following examples illustrate how various forms of data visualization, such as hierarchical trees, network graphs, and radial diagrams, have been integrated within the scope of this thesis to enhance the systematic classification and comprehension of digital architecture knowledge.

Information visualization necessitates a series of preprocessing steps to prepare data for effective visual analysis. This process typically involves extracting data from various sources, transforming and enriching it with supplementary metadata, and structuring it into appropriate models. Figure 4.2 illustrates an abstract representation of this workflow, outlining a comprehensive, user-centered, end-to-end roadmap that encapsulates the key tasks, potential issues, and challenges associated with data visualization in general (Po et al, 2020).

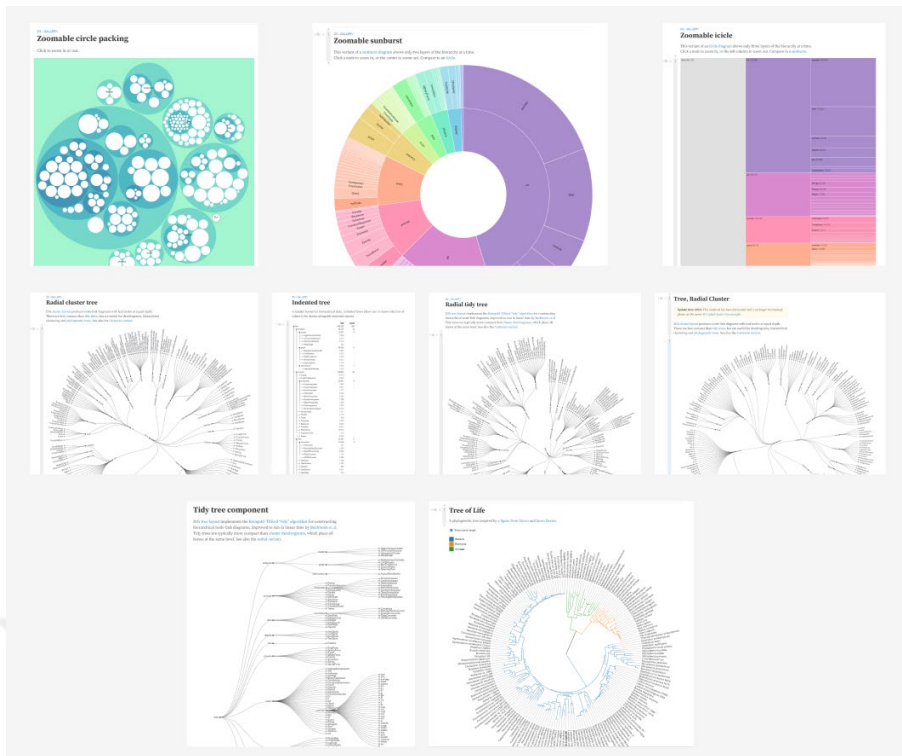


Figure 4.1 : Data visualization examples of D3.js (Url-8).

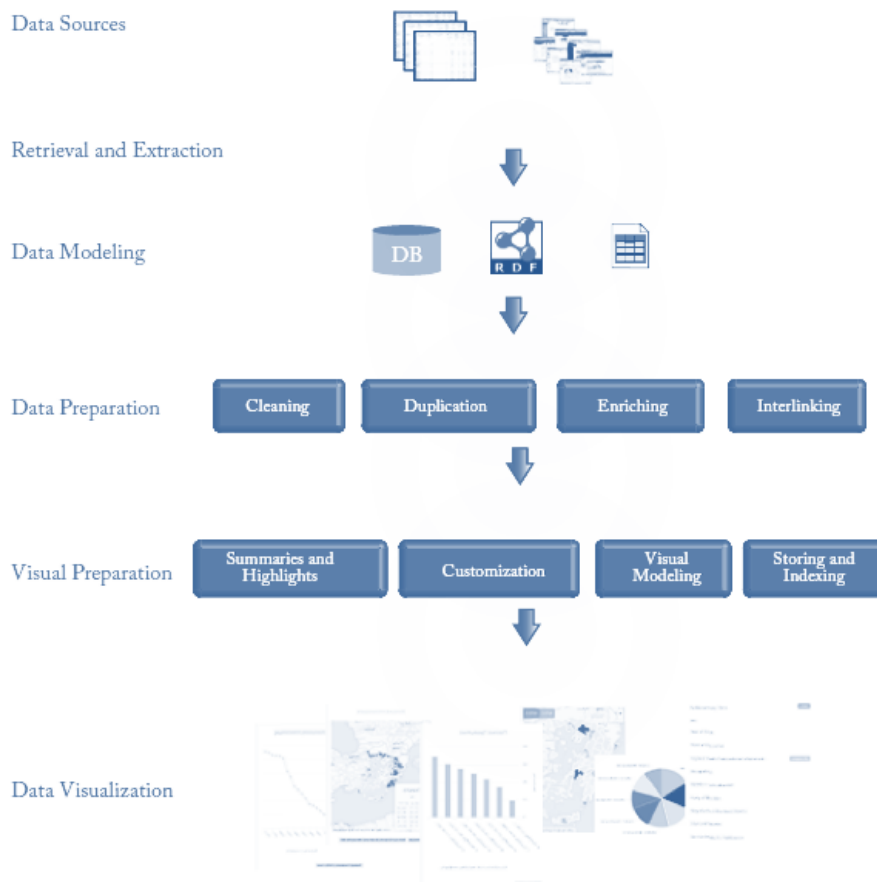


Figure 4.2 : The flowchart of the preprocessing steps and data visualization (Po et. al., 2020).

4.3 Phase 1: Testing Text Classification and Hierarchical Visualization with a Sampled Dataset

The first phase of the study uses an online encyclopedia for exploring hierarchical visualization across four knowledge layers: conceptual, procedural, visual, and configurational. Instead of presenting high-level information all at once, it facilitates phased knowledge transfer through structured layers. Data visualization tools are employed to guide users on unfamiliar topics and illustrate interrelationships among information visually.

The study methodology comprises five stages as depicted in Figure 4.3 and 4.4:

1. Retrieval of Raw Data from CUMINCAD Database
2. Preprocessing of the Data Set and Texts
3. Pre-categorization of the Data and Creation of an Excel File
4. Categorization of Articles
5. Hierarchical Visualization of Categorized Information with Six Diagram Types:

5-1. Zoomable, Circular, and Hierarchical Diagrams

5-2. Radial, Linear, and Hierarchical Diagrams

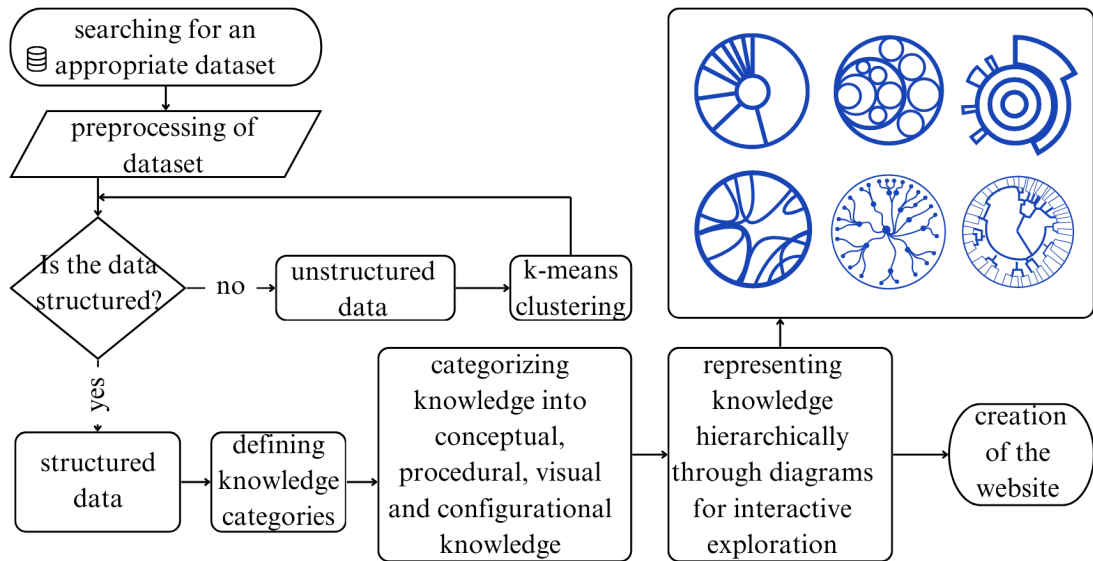


Figure 4.3 : The general flowchart of the methodology of the first phase with sampled data.

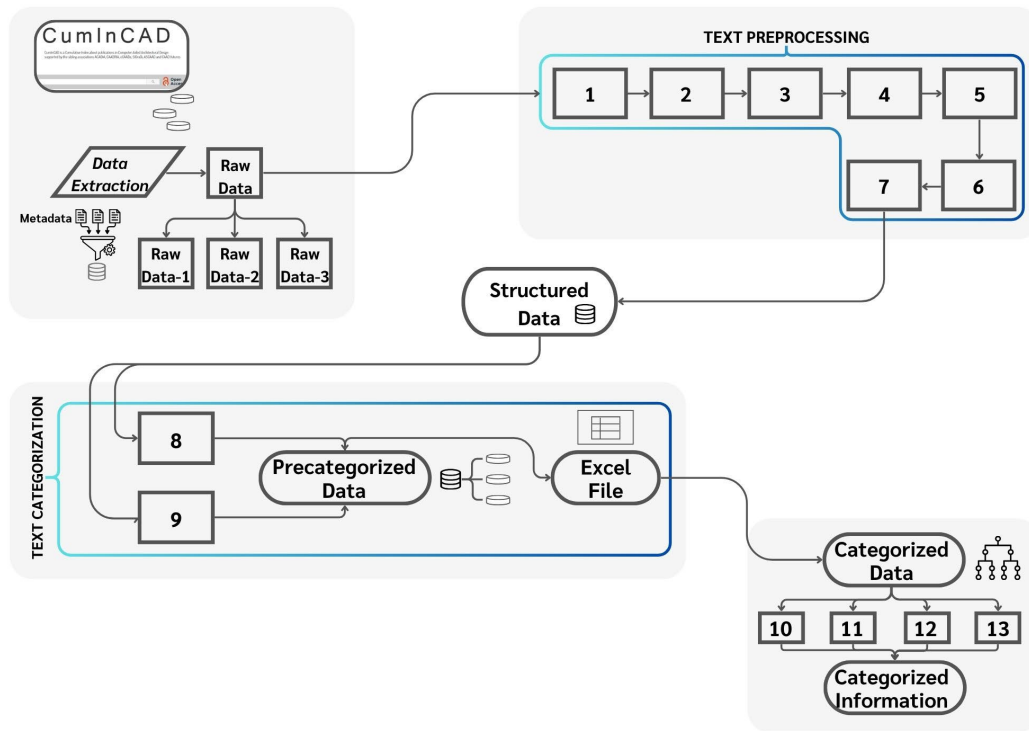


Figure 4.4 : The detailed flowchart of the methodology of the first phase with sampled data.

4.3.1 Retrieval of the raw data

The study employs the "Cumulative Index about Publications in Computer Aided Architectural Design (CUMINCAD)" database for categorizing information across the four knowledge layers. This database, containing over 12,300 bibliographic records from conferences such as ACADIA, ASCAAD, CAADRIA, eCAADe, SiGraDi, CAAD Futures, and DDSS, provides comprehensive abstracts and full-text PDFs for 9,600 articles (Url-9). Its significance lies in its extensive dataset and support for data formats like .ris, .bibtex, and Endnote. This resource aids in hierarchical clustering to uncover key themes, relationships, and patterns within digital architecture research. The system created within the scope of this research includes retrieving data from CUMINCAD database as in the Figure 4.5.

Figure 4.5 : Extracting data from CUMINCAD database (Url-9).

```

TY - ELEC
ID - e978
AU - [Zupancic] Strojan, Tadeja Z.
PY - 1999///
T1 - CyberUniversity
AD - Architectural Computing from Turing to 2000 [eCAADe Conference Proceedings / ISBN 0-9523
KW - University, Cyberuniversity, Space, Cyberspace
UR - http://papers.cumincad.org:443/cgi-bin/works/Show?
N2 - The study of a cyberuniversity derives from an analogy between real urban space and its
ER -

TY - ELEC
ID - ddss2006-hb-467
AU - A. Fatah gen. Schieck, A. Penn, V. Kostakos, E. O'Neill, T. Kindberg, D. Stanton Fraser,
PY - 2006///
T1 - Design Tools for Pervasive Computing in Urban Environments
AD - Van Leeuwen, J.P. and H.J.P. Timmermans (eds.) 2006, Innovations in Design & Decision Su
KW - Urban space, Pervasive systems, Urban computing, Space Syntax, Interaction space
UR - http://papers.cumincad.org:443/cgi-bin/works/Show?DDSS2006-HB-467
N2 - In this paper we report on ongoing research in which the implications of urban scale per
ER -

TY - ELEC
ID - ijac202018304
AU - Aagaard, Anders Kruse and Niels Martin Larsen
PY - 2020///
T1 - Developing a fabrication workflow for irregular sawlogs
AD - International Journal of Architectural Computing vol. 18 - no. 3, 270-283
KW - Natural wood, robotic fabrication, computation, fabrication, research by design
UR - http://papers.cumincad.org:443/cgi-bin/works/Show?ijac202018304
N2 - In this article, we suggest using contemporary manufacturing technologies to integrate m
ing/digitisation, handling of a stockpile, computer analysis, design and robotic manufacturing
ER -

TY - ELEC
ID - ecaade2014_021
AU - Aant van der Zee, Bauke de Vries and Theo Salet
PY - 2014///

```

Figure 4.6 : “.ris” file format retrieved from the CUMINCAD database.

The data downloaded from the CUMINCAD database is provided in a ".ris" file format (Figure 4.6), encompassing article titles, keywords, abstracts, authors, publication years, and conference names. Initially, titles, keywords and summaries were utilized, and the text file was converted into an Excel spreadsheet for categorization.

4.3.2 Preprocessing of the data set

The first phase of this study focuses on digital fabrication subfield within digital architecture, using metadata from the CUMINCAD platform, including titles, keywords, abstracts, authors, publication years, and conference names. The dataset

was refined by selecting titles (Raw Data-1), keywords (Raw Data-2), and abstracts (Raw Data-3) of the articles.

Preprocessing, conducted with Python, involved:

1. Removal of Rows with Cells Containing Non-English Words
2. Removal of Columns with Keywords Containing Non-English Words
3. Capitalization of Each Word
4. Correction of Spelling Mistakes
5. Removal of Rows with Empty Cells
6. Extraction of File with Separated Keywords and Removal of Cells with More Than 5 Words
7. Lowercasing of Certain Common Words, Excluding Those Found at the Beginning of a Title (of, on, and, to, a, in, for, by)

These steps are carried out using Python for back-end development and HTML, CSS, and JavaScript for front-end implementation, with the support of generative AI tools such as ChatGPT and Google AI Studio in code generation. The cleaned data was then formatted for categorization (Figure 4.7)

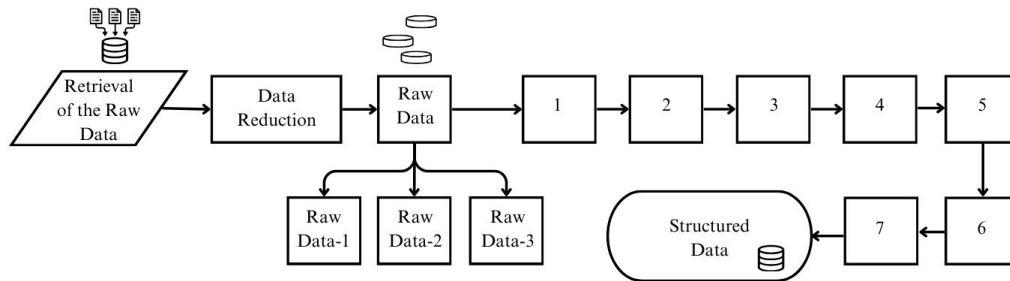


Figure 4.7 : Raw data extraction from CUMINCAD and transformation into structured data after reduction and preprocessing.

It is important to note the level of manual intervention in this process. While the goal was maximum automation, a fully automated pipeline presented challenges to semantic accuracy. For instance, algorithmic keyword clustering occasionally grouped conceptually distinct terms (e.g., 'model' as a physical object vs. 'model' as a computational process). Therefore, a manual review and refinement stage was implemented to validate the clusters and ensure the logical integrity of the knowledge hierarchy. This human-in-the-loop approach was a necessary step to overcome the limitations of current NLP technologies in discerning nuanced,

domain-specific context, ensuring the final classified data was both computationally structured and academically sound.

4.3.3 Categorization of the preprocessed data set

Unsupervised learning techniques, such as the k-Means algorithm, uncover latent connections and patterns in extensive digital architecture article datasets. This algorithm groups articles based on similarities in titles and keywords without needing predefined labels, thus automating and streamlining the classification process. Given the laborious nature of manual organization, k-Means significantly reduces time and effort by clustering articles into predefined numbers of groups. This enables students and researchers to better understand research domains and trends within digital architecture.

Before applying the k-Means algorithm, the TF-IDF vectorizer is utilized to convert text data into a numerical format, evaluating the weights of words within documents. This process helps differentiate topics based on cluster-term distribution. TF-IDF (Number 8 in Figure 4.8) combines Term Frequency (TF) and Inverse Document Frequency (IDF), with IDF calculated as the logarithm of the ratio of total documents to the number containing the term (Grootendorst, 2022). The k-Means algorithm (Number 9 in Figure 4.8) groups articles based on these numerical representations, automating classification and streamlining the organization of extensive datasets (Figure 4.8).

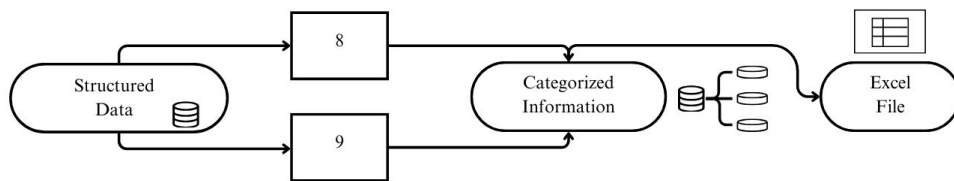


Figure 4.8 : Diagram illustrating categorization of structured data using TF-IDF and k-Means clustering, followed by export to Excel.

The hierarchical text classification output related to the theme of Fabrication demonstrates the nuanced and multidimensional structure of knowledge within digital architecture framework. At the top level, the main cluster is identified as Fabrication, which is further divided into several subclusters and sub-subclusters reflecting diverse technological and methodological approaches. The Robotic Fabrication subcluster, for instance, encompasses themes such as Collaborative

Robotic Fabrication, Interactive Robotic Fabrication, Precision Robotic Fabrication, and Evolutionary Fabrication, each emphasizing distinct modes of robotic interaction and automation. The Digital Fabrication subcluster includes categories like Design and Digital Fabrication, Full-Scale Digital Fabrication, Computational Fabrication, and Additive Manufacturing, highlighting the integration of digital design tools, scale-sensitive construction techniques and material optimization strategies. Additional subclusters such as Other Fabrication Techniques, Specialized Fabrication and Intersection of Digital and Robotic Fabrication cover themes including Adaptive and Augmented Fabrication, CNC-based Fabrication, Textile and Bio-Fabrication and Hybrid Fabrication. This layered structure not only reveals the thematic richness within fabrication-related studies but also supports a more granular exploration of emerging subfields and intersections within digital and robotic processes, offering a comprehensive framework for knowledge mapping and trend analysis in computational design research.

In this particular context, the data visualizations are presented uniformly. This code facilitates reusability through the utilization of a consistent JSON document format. This JSON file allows to reuse it in D3.js with different data visualization examples. The format of a standard .json file itself is hierarchical.

In the final stage of the study, it is suggested to develop an exploratory data visualization model utilizing D3.js using generative AI to generate codes. This model visually represents relationships using hierarchical diagrams such as "zoomable sunburst, zoomable circle packing, sequences sunburst, radial tree, hierarchical edge bundling and tree of life". This model enables users to explore clustered and classified information discovery and hierarchical exploration through various graphics, allowing them to analyze sub-relationships among connections.

4.3.4 Visualization of the categorized data: Comparative analysis of hierarchical diagrams

In the final stage of the first phase of this study, an exploratory data visualization model (Number 10) using D3.js is proposed. This model employs hierarchical diagrams, including "Zoomable Sunburst (Number 11-1), Zoomable Circle Packing (Number 11-2), Sequences Sunburst (Number 11-3) (First Group-Number 11 in Figure 4.9); Hierarchical Edge Bundling (Number 12-1), Radial Tree (Number 12-2),

Tree of Life (Number 12-3) (Second Group-Number 12 in Figure 4.9)". These visualizations allow users to explore clustered and classified information through various graphs, enabling the discovery of connections and the analysis of sub-relationships among the data (Number 13 in Figure 4.9). The first group features zoomable, circular, solid, and hierarchical representations, while the second group includes radial, linear, and hierarchical structures (Figure 4.9).

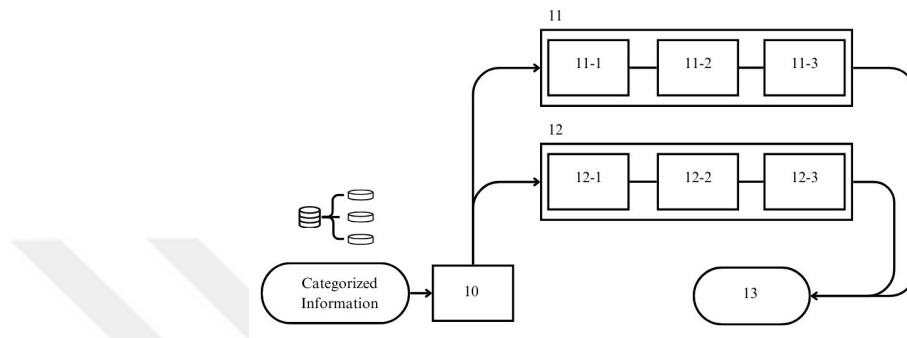


Figure 4.9 : Visualization of categorized data and model building using two sets of data visualization tools.

In the resulting online encyclopedia of digital architecture, circular forms predominate, both in the logo and diagrams. The circular shape is employed for its symbolic ability to focus attention on a central point, facilitate sequential navigation in hierarchical data, and ensure ease and fluidity when moving between categories and subcategories. It also supports smooth transitions, enhances memorability and symbolizes the continuous loop and interconnectedness of information. Furthermore, it reflects the ongoing updating, integrity, multilayered nature, and ambition to serve as a comprehensive information source. Circular and radial graphs have been chosen for their aesthetic coherence and consistency, with the center representing the main topic or core information and the surrounding rings indicating subtopics or related data.

The selection of diagrams for information retrieval is left to the user's discretion. Therefore, the advantages and disadvantages of each diagram from the user's perspective are outlined in Table 4.1 and Table 4.2.

4.3.4.1 First group: Zoomable, radial, solid and hierarchical representation

These diagrams have been examined under three groups. The Zoomable Sunburst Diagram is a radial hierarchical representation. Only the first two layers of the

hierarchy structure are displayed simultaneously. The Sequences Sunburst Diagram transfers data from the Cartesian axis to the polar axis, allowing the graph to be read through angles. The Zoomable Circle Packing Diagram consists of nested circles that are hierarchically organized in a circular form (Figure 4.10).

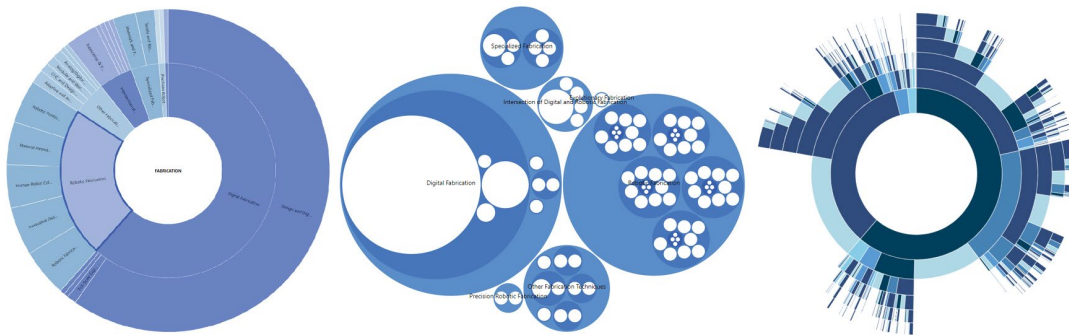


Figure 4.10 : Zoomable, circular, solid, and hierarchical diagrammatic representations (From left to right: zoomable sunburst, zoomable circle packing, sequences sunburst).

Table 4.1 : Descriptions, advantages and disadvantages of first group of diagrams.

Diagram Type	Description	Advantages	Disadvantages
Zoomable Sunburst	A radial hierarchical visualization that expands progressively from the center outward.	-The initial complexity of data is reduced by displaying only the top two hierarchical layers simultaneously. -Facilitates understanding of relationships between data through zooming in and out. -Offers a compact representation with an angular depiction.	-As the diagram's content expands, there is a risk of reduced readability within the circle.

Table 4.2 (continued) : Descriptions, advantages and disadvantages of first group of diagrams.

Diagram Type	Description	Advantages	Disadvantages
Zoomable Circle Packing	Composed of nested clusters in a circular form, arranged hierarchically.	-Allows access to various regions and clusters through zoomable animations. -Enables users to quickly evaluate specific clusters and associated data.	-Due to the nested nature of clusters, deeper layers may become harder to interpret, especially as the diagram grows, leading to potential difficulties in quickly identifying and accessing deeper information.
Sequences Sunburst	A graphic mapping linear data on a Cartesian axis to a polar axis, allowing single-instance readings.	-Provides clear visibility of data at the greatest depth. -Allows rapid access to data in the lowest layer.	- Readability issues may arise in complex datasets. -Readability of the last data member may be low.

4.3.4.2 Second group: Radial, linear, network-based and hierarchical representation

These diagrams have been examined under three groups. The Radial Tree Diagram is a graphical representation that visualizes hierarchical data in a tree diagram format, allowing for the expansion and collapse of nodes. The Hierarchical Edge Bundling Diagram allows for the visualization of adjacency relationships among entities organized in a hierarchy. The idea here is to group adjacent edges together to reduce the clutter typically observed in complex networks. The Tree of Life diagram, also known as the Phylogenetic Tree, is a visual representation that illustrates the evolutionary relationships between different species or groups of organisms (Figure 4.11).

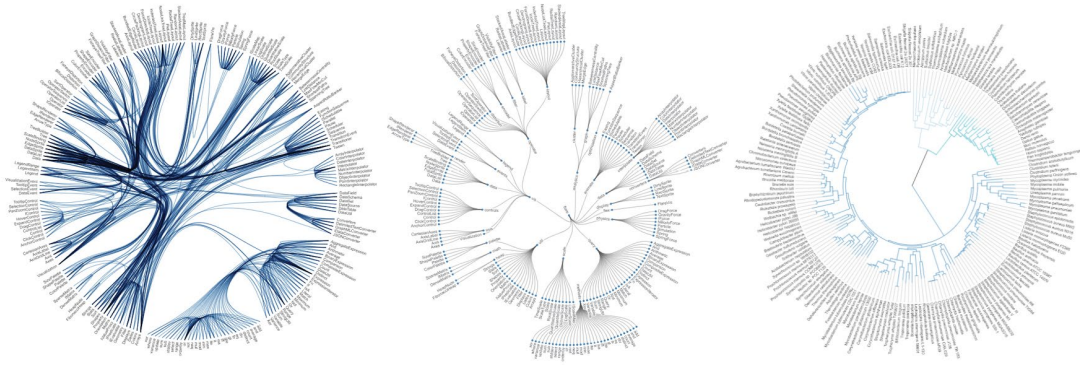


Figure 4.11 : Radial, linear, and hierarchical diagrammatic representations (From left to right: hierarchical edge bundling, radial tree, tree of life).

Table 4.3 : Descriptions, advantages and disadvantages of second group of diagrams.

Diagram Type	Description	Advantages	Disadvantages
Hierarchical Edge Bundling	Facilitates the visualization of adjacency relationships within hierarchical structures.	-Provides comprehensive network analysis by visualizing both direct and implicit adjacency relationships between parent nodes.	-During edge bundling, some details or relationships may be visually obscured.
Radial Tree	Used to create hierarchical node-link diagrams in a tree structure.	-Displays hierarchical structures in a compact, layered format.	-Node depth calculation relative to the root may cause irregular appearance and reduced readability.

Table 4.2 (continued): Descriptions, advantages and disadvantages of second group of diagrams.

Diagram Type	Description	Advantages	Disadvantages
Tree of Life	A phylogenetic tree, resembling a circular dendrogram, inspired by tree branches.	-Clearly depicts evolutionary relationships and processes, enhancing the comprehensibility of complex data.	-An excess of metadata can complicate the visualization, making accurate data interpretation difficult.

4.3.5 Web interface for the classification and visualization of digital architecture knowledge

As part of the proposed framework, an interactive web interface was developed to facilitate the classification and visualization of digital architecture knowledge. This interface integrates zoomable sunburst diagrams that dynamically represent hierarchical knowledge structures across four distinct layers: conceptual, procedural, visual, and configurational. These layers are derived from the content of scholarly articles and serve to visually communicate the multidimensional nature of digital architecture knowledge.

One prominent example is the Zoomable Sunburst Diagram in the context of the Web Page (Figure 4.12), which offers a circular, layered visualization that allows users to intuitively explore the distribution and relationships of knowledge categories within a selected article. As shown in the Zoomable Sunburst Diagram in the Web Page (Figure 4.13), users can interactively zoom into subcategories and examine the underlying knowledge structure in greater detail. The diagram serves as both a navigational and analytical tool within the interface.

The web interface showcases specific articles to demonstrate this functionality. For instance, the article '*Collaborative Robotic Masonry and Early Stage Fatigue Prediction*' (Das et al., 2019) was selected from the dataset without bias for visualization. This selection strategy was adopted to avoid bias toward highly cited or thematically dominant publications, while also ensuring that the chosen study holds sufficient representational value in terms of content diversity and relevance. Its

corresponding Conceptual (a), Procedural (b), Visual (c), and Configurational Knowledge (d) layers are displayed on the right side of the webpage (Figure 4.14). Simultaneously, the full-text PDF of the article is embedded within an iframe on the left, allowing users to read the document alongside its visualized knowledge structure. This dual presentation not only enhances comprehension but also encourages users to critically engage with the classification process.

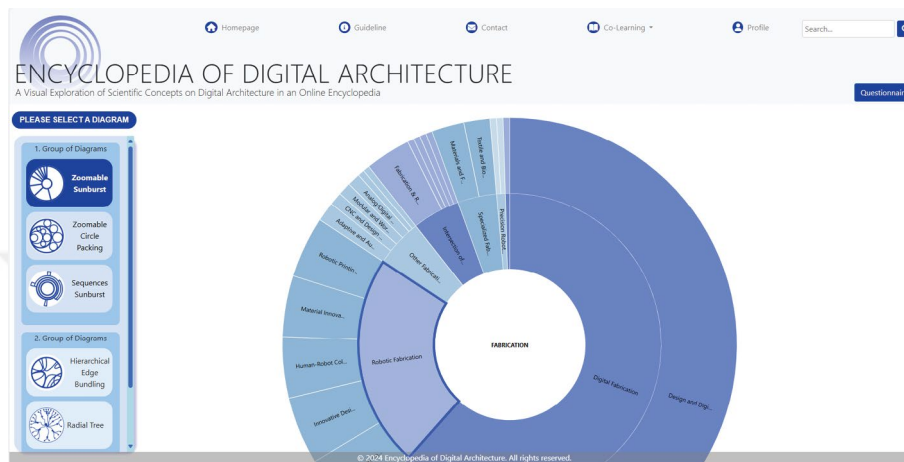


Figure 4.12 : Zoomable sunburst diagram in the context of the web page.

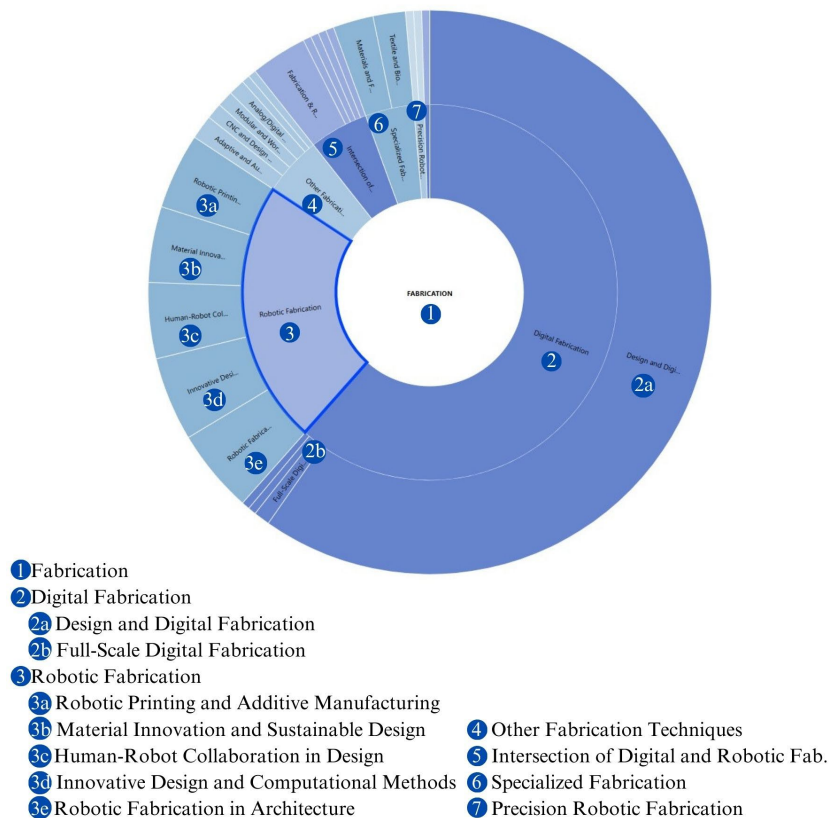


Figure 4.13 : The elements of zoomable sunburst diagram in the web page.

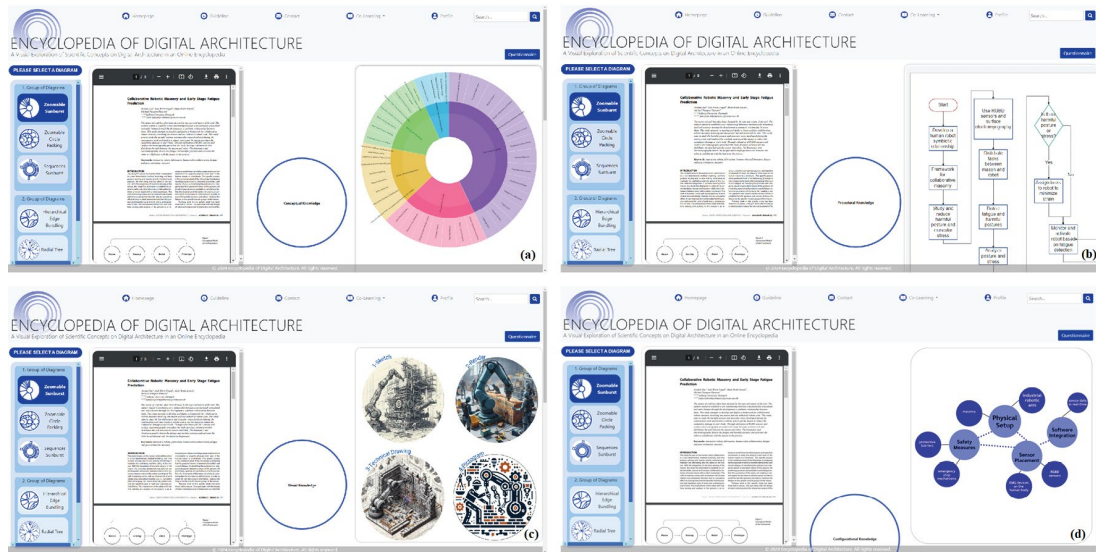


Figure 4.14 : Conceptual (a), procedural (b), visual (c), and configurational knowledge (d) layer of the article (on the right side of the webpage) and the PDF of the article in an iframe on the webpage (left).

Furthermore, the Subcategories of Robotic Fabrication (Left) and Article with Four Knowledge Layers (Right) section of the interface (Figure 4.15) allows users to explore related subdomains and compare knowledge structures across different publications. Inside the Zoomable Sunburst Diagram further exemplifies how complex content can be visually decomposed and hierarchically organized within the web environment (Figure 4.16).

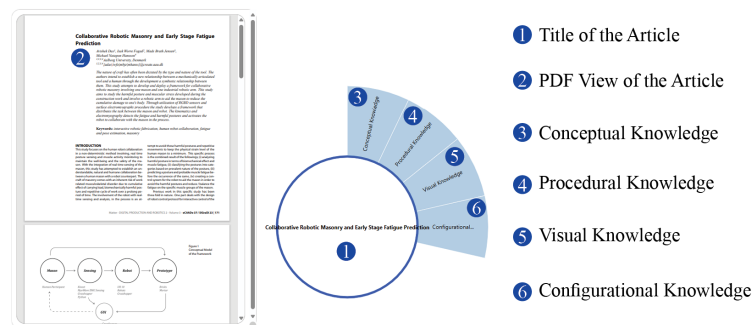


Figure 4.15 : Subcategories of Robotic Fabrication (left) and article with four knowledge layers (right).

Overall, this interface exemplifies how interactive data visualization (particularly through D3.js-based sunburst diagrams) can support the classification, representation, and navigation of digital architecture knowledge, offering users an engaging and analytically rich platform for exploratory learning.

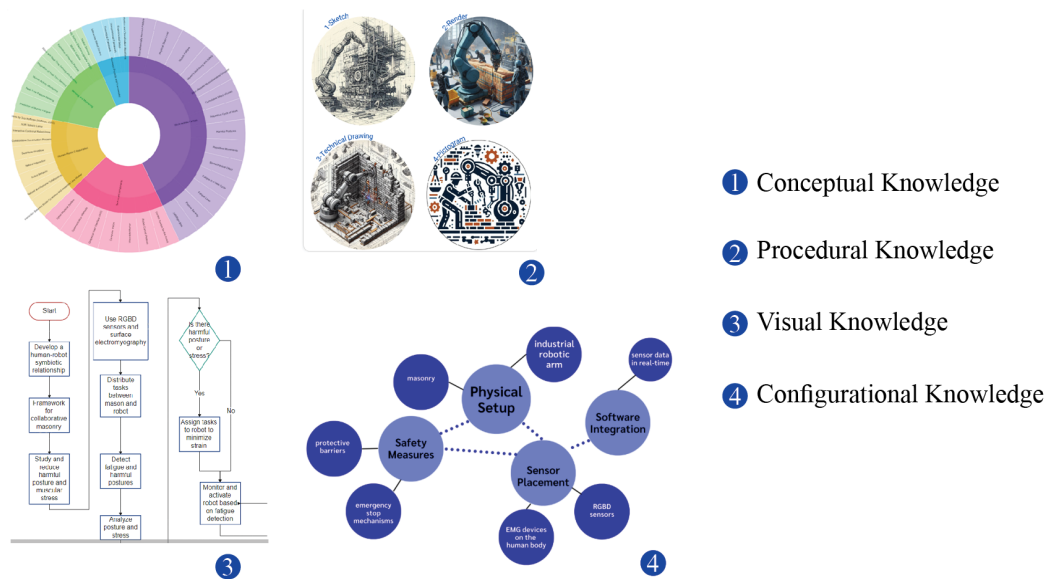


Figure 4.16 : Conceptual, procedural, visual, and configurational knowledge layer of the article “Collaborative Robotic Masonry and Early Stage Fatigue Prediction” inside the zoomable sunburst diagram.

4.4 Phase 2: Testing Text Classification and Hierarchical Visualization with a Full Dataset

The following sections present a systematic overview of the core components of the web interface designed to support the processing, classification and visualization of a comprehensive dataset derived from the field of digital architecture. This phase focuses on evaluating the effectiveness and scalability of the proposed classification framework when applied to a large corpus of domain-specific texts. Each interactive element, ranging from file upload mechanisms to data visualization modules, is described in terms of its function and purpose, illustrating how the system facilitates user interaction, manages input data and enables the exploration of hierarchical relationships within classified knowledge structures. Collectively, these components constitute an integrated prototype that operationalizes both the logic of text classification and the visual representation of multi-layered thematic clusters.

4.4.1 Creation of RIS to JSON converter and categorized hierarchical keyword classification plug-ins

In the second phase of the study, a custom plug-in was developed and integrated into the homepage as an embedded framework to facilitate the conversion of retrieved RIS-formatted bibliographic data into JSON format. This tool streamlines data

processing by enabling seamless transformation of complex RIS metadata into a structured and machine-readable JSON format, optimizing its usability for subsequent analysis. Complementing this functionality, an advanced feature, referred to as the JSON Hierarchical Clustering tool, was implemented to organize the converted JSON data into a hierarchical structure. This clustering mechanism leverages computational methods to categorize and group the data based on predefined criteria, thereby establishing a coherent and systematic arrangement of the information. Together, these tools enhance the efficiency of data management and analysis, enabling the structured representation and exploration of large datasets within the digital architecture knowledge framework (Figure 4.17).

The following framework outlines the procedural stages implemented by the author, subsequently refined with the assistance of generative AI (ChatGPT), for the design and development of a .RIS to JSON conversion tool. The process begins with the setup of a web-based interface comprising clearly structured components: a user instruction section, a drag-and-drop area alongside a traditional file input for .RIS file uploads, dedicated buttons for initiating data conversion and downloading outputs, as well as designated areas for displaying the resulting JSON and statistical summaries (e.g., total, eliminated and retained articles).

Event listeners are then integrated into the interface to facilitate interactive functionality. These listeners handle user interactions such as clicking the drop zone to activate file selection, managing the drag-and-drop behavior for visual responsiveness, and detecting changes in the file input to trigger file processing routines.

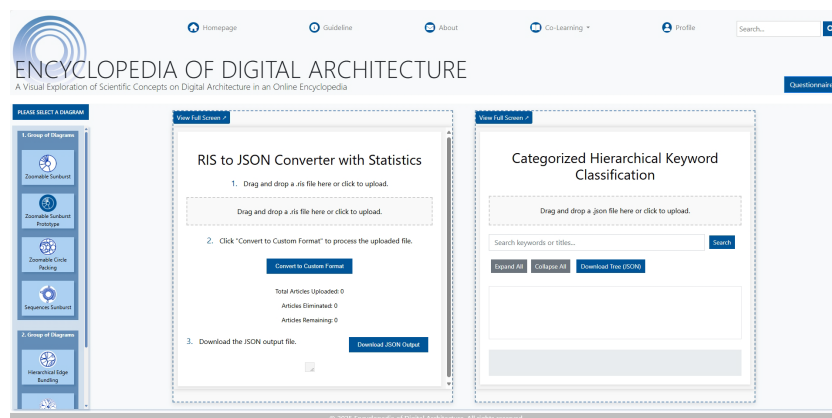


Figure 4.17 : RIS to JSON converter and hierarchical clustering of JSON data on the homepage.

Once a file is uploaded, it undergoes a preliminary validation phase to confirm its .RIS format. Upon validation, the system provides a visual cue indicating that the file is being loaded, and its contents are read and passed into a dedicated parsing function. The parsing mechanism processes the raw .RIS data line by line, identifying entries based on the TAG - VALUE syntax and grouping them into discrete articles using the TY tag as a delimiter. These grouped articles are stored in an array for further manipulation.

Upon user initiation via the "Convert to Custom Format" button, the application proceeds to process the parsed data. Each entry undergoes normalization and transformation, including the sanitization of non-ASCII characters, the reformatting of keywords into standardized uppercase comma-separated strings and the evaluation of summaries to ensure sufficient detail, typically exceeding 100 words. Articles failing to meet these criteria are filtered out, and statistics regarding the total number of processed articles, those eliminated and those retained are calculated and displayed. The remaining articles are then converted into a structured JSON format, mapped with custom keys such as "Author(s)", "Year", "Title", "Keywords" and "Summary" (Appendix E).

A download mechanism is provided through a designated button, which, when clicked, verifies the availability of transformed data, generates a JSON file named with the current date, and automatically triggers the download process. Throughout the interaction, the interface offers real-time visual feedback, highlighting the drop zone during drag operations and updating status messages to reflect stages such as loading or successful upload, thus enhancing the overall user experience. This structured and automated workflow ensures a robust, user-friendly and scalable approach for converting and analyzing bibliographic data from .RIS format to a machine-readable JSON structure.

In the process of transforming a .RIS file into a structured JSON format, specific bibliographic tags are systematically mapped to custom-labeled fields to facilitate data readability and usability. The AU tag, which indicates the author's name, is assigned to the "Author(s)" field in the JSON output. The PY tag, representing the publication year, is cleaned by removing any trailing characters (e.g., slashes) and mapped to the "Year" field. The primary title of the work, marked by the T1 tag in the .ris file, corresponds to the "Title" field in JSON. Keywords listed under the KW

tag are normalized, typically converted to uppercase and separated by commas and stored in the "Keywords" field. The abstract or summary of the document, provided under the N2 tag, is directly transferred to the "Summary" field after optional formatting adjustments. Other tags such as TY (type of reference), ID, AD (conference location and citation), and UR (URL) are omitted, as they are not required in the customized output. This transformation ensures that the essential content of each entry is retained in a clear, consistent, and machine-readable format suitable for further processing or analysis (Figure 4.18).

RIS to JSON Converter with Statistics

1. Drag and drop a .ris file here or click to upload.

Drag and drop a .ris file here or click to upload.
2. Click "Convert to Custom Format" to process the uploaded file.

Convert to Custom Format

Total Articles Uploaded: 10100
 Articles Eliminated: 4086
 Articles Remaining: 6014
3. Download the JSON output file.

Download JSON Output

```
[
  {
    "Author(s)": "[Zupancic] Strojan, Tadeja Z.",
    "Year": "1999",
    "Title": "CyberUniversity",
    "Keywords": "UNIVERSITY, CYBERUNIVERSITY, SPACE, CYBERSPACE",
    "Summary": "The study of a cyberuniversity derives from an analogy between real urban space and its virtual \\'substitution\\'. It represents an attempt to balance some views, which seems to be contrary, exclusive, but they are just parts of the same wholeness. Especially the notion of a cyber society is lately considered such an exaggeration, that it is possible to forget the meaning of a real life experience and interactions, which are already threatened. One should contribute to the awareness that is used in such a comparison, it is \\'just\\' an analogy, not a real similarity. At the same time it is possible to point out some limitations of a cyberspace and indicate a more realistic view of the meaning of cyber communities. Awareness of the development processes could help to find a balance between reality and virtuality, using cyberfacilities not to destroy us (our identity) but to improve the quality of our (real) life."
  },
  {
    "Author(s)": "A. Fatah gen. Schieck, A. Penn, V. Kostakos, E. O'Neill, T. Kindberg, D. Stanton Fraser, and T. Jones",
    "Year": "2006",
    "Title": "Design Tools for Pervasive Computing in Urban Environments",
    "Keywords": "URBAN SPACE, PERSVASIVE SYSTEMS, URBAN COMPUTING, SPACE SYNTAX, INTERACTION SPACE",
    "Summary": "In this paper we report on ongoing research in which the implications of urban scale pervasive computing (always and everywhere present) are investigated for urban life and urban design in the heritage environment of the city of Bath. We explore a theoretical framework for
```

The file is uploaded.

Figure 4.18 : The output of the RIS to JSON converter on the homepage when a .ris file is uploaded.

An example of the .ris file input:

TY - ELEC

ID - e978

AU - [Zupancic] Strojan, Tadeja Z.

PY - 1999///

TI - CyberUniversity

AD - Architectural Computing from Turing to 2000 [eCAADe Conference Proceedings / ISBN 0-9523687-5-7] Liverpool (UK) 15-17 September 1999, pp. 196-200

KW - University, Cyberuniversity, Space, Cyberspace

UR - <http://papers.cumincad.org:443/cgi-bin/works/Show?>

N2 - *The study of a cyberuniversity derives from an analogy between real urban space and its virtual "substitution". It represents an attempt to balance some views, which seems to be contrary, exclusive, but they are just parts of the same wholeness. Especially the notion of a cyber society is lately considered such an exaggeration, that it is possible to forget the meaning of a real life experience and interactions, which are already threatened. One should contribute to the awarness that is used in such a comparison, it is "just" an analogy, not a real similarity. At the same time it is possible to point out some limitations of a cyberspace and indicate a more realistic view of the meaning of cyber communities. Awarness of the development processes could help to find a balance between reality and virtuality, using cyberfacilities not to destroy us (our identity) but to improve the quality of our (real) life.*

ER -

Here is the output of the .ris file as a json file:

```
{
  "Author(s)": "[Zupancic] Strojan, Tadeja Z.",
  "Year": "1999",
  "Title": "CyberUniversity",
  "Keywords": "UNIVERSITY, CYBERUNIVERSITY, SPACE, CYBERSPACE",
  "Summary": "The study of a cyberuniversity derives from an analogy between real urban space and its virtual \"substitution\". It represents an attempt to balance some views, which seems to be contrary, exclusive, but they are just parts of the same wholeness. Especially the notion of a cyber society is lately considered such an exaggeration, that it is possible to forget the meaning of a real life experience and interactions, which are already threatened. One should contribute to the awarness that is used in such a comparison, it is \"just\" an analogy, not a real similarity. At the same time it is possible to point out some limitations of a cyberspace and indicate a more realistic view of the meaning of cyber communities. Awarness of the development processes could help to find a balance between reality and virtuality, using cyberfacilities not to destroy us (our identity) but to improve the quality of our (real) life."
}
```

Thematic analysis and knowledge structure identification within extensive research corpora present significant challenges in bibliometric studies. Figure 4.19 introduces a novel web-based system for Categorized Hierarchical Keyword Classification, engineered to process scholarly metadata (titles, keywords, summaries) from JSON inputs and generate a meaningful, navigable hierarchy. The system employs a multi-stage pipeline designed to transform raw keywords into a structured representation that enhances user understanding and exploration.

The core processing stages are as follows: Initially, publication data, including titles and keywords, are ingested from the JSON file. Keywords are extracted and standardized. This involves converting keywords to lowercase, trimming extraneous

whitespace, and selectively handling hyphenated terms. Certain predefined hyphenated terms are preserved intact, while other hyphens are converted to spaces to treat components individually. Subsequently, all processed keywords are converted to uppercase for consistent matching and stored uniquely per publication. Concurrently, publication details are cached for subsequent retrieval.

Next, a hierarchical structure is constructed from the processed keywords. Multi-word phrases form branching tree structures, where each word constitutes a node, and the publication's title is associated with the terminal node of the phrase. This initial hierarchy is then refined through synonym consolidation, which utilizes a predefined synonym map and a simple stemming algorithm to group semantically equivalent or related terms under a single canonical form.

Following this, the consolidated keyword nodes are subjected to multi-level categorization. They are first assigned to broad thematic main categories using predefined mappings and secondary matching rules. Keywords within each main category are further allocated to more granular sub-categories based on configured term lists. Keywords that do not map to any defined sub-category within a main category are assigned to a default miscellaneous sub-category. If a keyword does not fit any main category, it is routed to a top-level miscellaneous category, which itself has its own sub-categorization logic, with any remaining unassigned terms falling into a final uncategorized group.

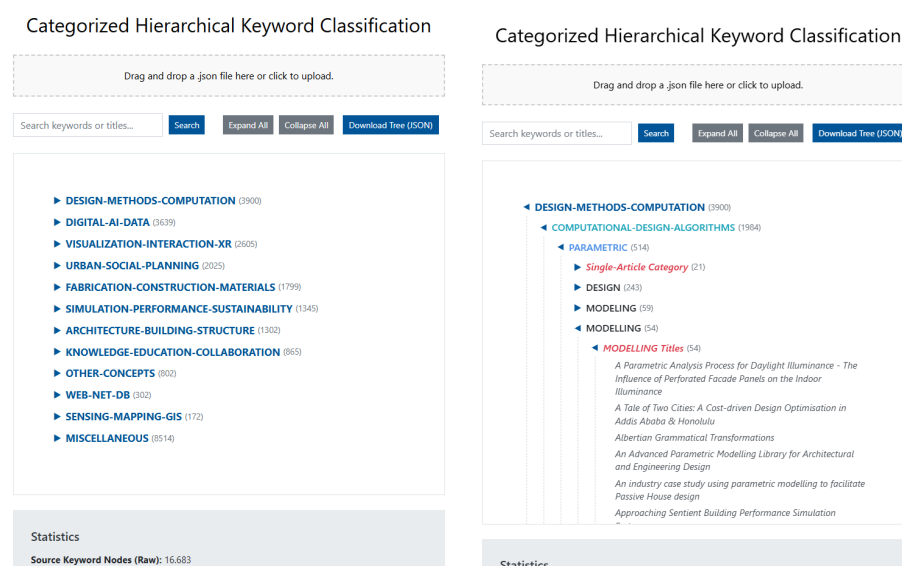


Figure 4.19 : The output of the categorized hierarchical clustering of JSON data on the homepage.

To further enhance structural coherence and semantic meaning, the system incorporates several refinements. For instance, a specific process targets bio-related terms within the materials science domain, consolidating all terms prefixed with "BIO-", the keyword "BIO" itself, and their respective child nodes and associated titles under a new, singular "BIO" group node. Another process identifies keywords associated with only a single publication and no further child keywords; if the count of such nodes exceeds a defined threshold, they are aggregated into a 'Single-Article Category'. Additionally, to manage visual complexity, if the number of titles linked to a keyword exceeds a certain limit, or if a keyword has both sub-nodes and titles, these titles are grouped under a specially named node (e.g., "{Keyword Name} Titles") (Appendix F).

Finally, this comprehensively processed data structure is rendered as an interactive, collapsible HTML tree diagram. This visualization displays aggregated publication counts for each node and lists publication titles at the leaf level. Users can search the tree, expand/collapse branches, and view summary statistics regarding the generated structure. The entire hierarchical structure, including publication details as leaf nodes, can be exported in JSON format. This hierarchical clustered data is used to build the Zoomable Sunburst diagram with D3.js.

This approach offers a significant improvement in the discoverability of research literature, the elucidation of thematic patterns, and overall knowledge management, particularly in multidisciplinary fields such as digital architecture.

4.4.2 Zoomable sunburst diagram and prototype of the data visualization

In this section, the prototype of the data visualization developed within the scope of the study is presented, with a focus on the Zoomable Sunburst Diagram (Figure 4.20). This interactive diagram enables hierarchical representation of classified knowledge categories derived from the CUMINCAD dataset, facilitating intuitive navigation across multiple layers of information, namely, conceptual, procedural, visual, and configurational knowledge. Implemented using D3.js, the sunburst diagram offers smooth zooming transitions, hover-based tooltips and dynamic resizing, allowing users to explore complex relationships within the data without losing the structural context. Each arc segment in the diagram represents a category or subcategory, where segment size is proportional to the quantity or relevance of

entries. The zoomable interface enhances both macro-level understanding and micro-level analysis, aligning with the "overview first, zoom and filter, then details-on-demand" paradigm in information visualization. This interactive approach not only improves usability and clarity but also exemplifies how visual analytics can support knowledge acquisition in digital architecture research.

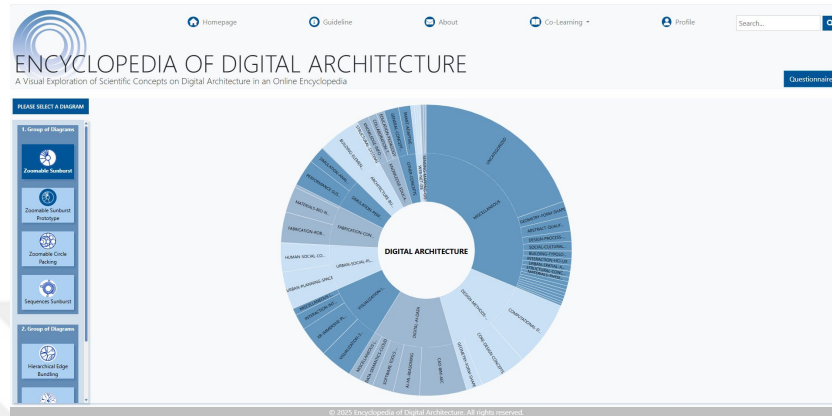


Figure 4.20 : Zoomable Sunburst diagram created with D3.js for data visualization.

The hierarchical clustered data is used to build the prototype of Zoomable Sunburst diagram with D3.js (Figure 4.21). The diagram illustrates three articles categorized under the main cluster URBAN-SOCIAL-PLANNING, within the subcluster HUMAN-SOCIAL-COGNITIVE-PARTICIPATION, and more specifically under the sub-subcluster HUMAN-ROBOT-COLLABORATION. Among these, the article titled “Collaborative Robotic Masonry and Early Stage Fatigue Prediction” is selected for detailed analysis. The subsequent figures demonstrate the classification of this article across four distinct layers of knowledge: conceptual, procedural, visual, and configurational.

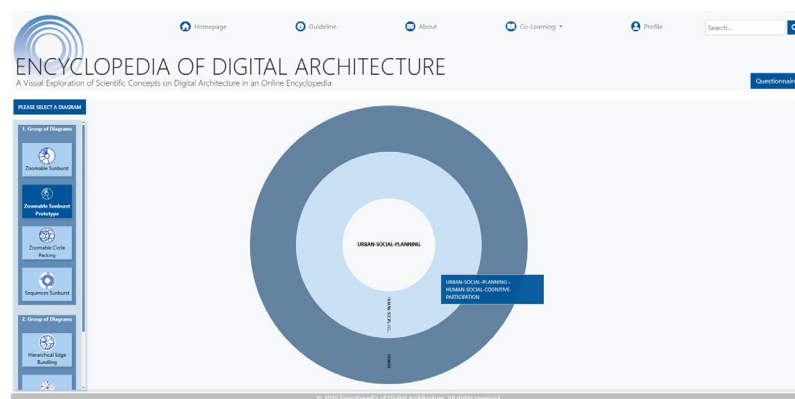


Figure 4.21 : Zoomable Sunburst Prototype diagram created with D3.js for data visualization.

presenting Procedural Knowledge in a manner that enhances both accessibility and comprehension (Figure 4.23).

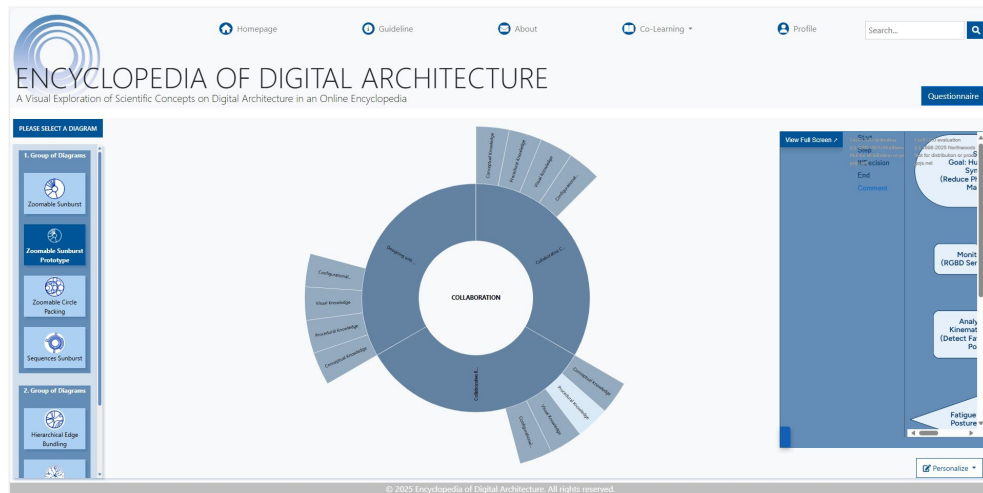


Figure 4.23 : The “Procedural Knowledge” represented by a flowchart.

The Visual Knowledge component is generated through Microsoft Designer. Specifically, four distinct types of visual representations were created for the user: a render, a technical drawing, a sketch, and a pictogram, each embodying a different structural and stylistic approach. These varied visual formats are intended to support the user in reinforcing their Visual Knowledge by offering multiple perspectives and interpretive layers of the same concept. Through this multimodal visual engagement, users can gain a deeper understanding of the article and its representational diversity (Figure 4.24).

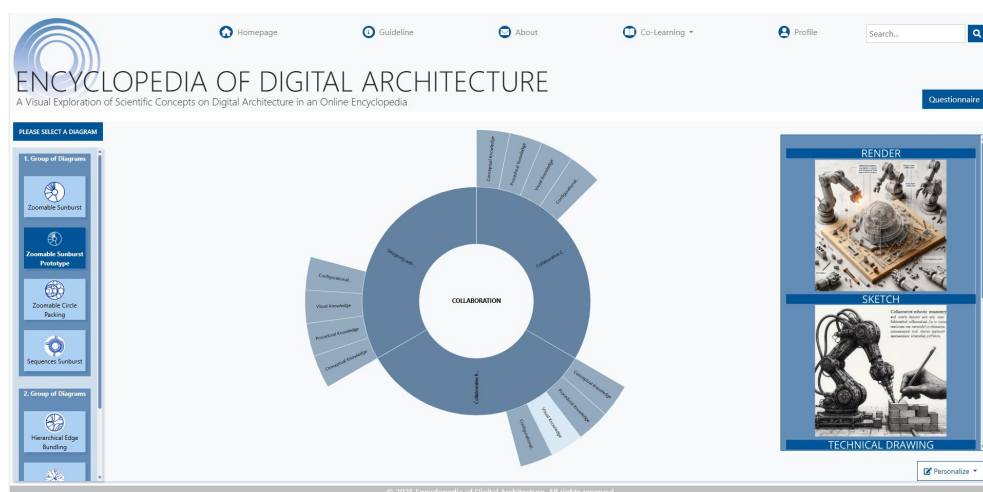


Figure 4.24 : The “Visual Knowledge” represented by Microsoft Designer.

The Configurational Knowledge component is represented through an interactive force-directed diagram, generated with the assistance of the GoJS library. This diagram visualizes the relationships and interconnections between various themes found within the articles' summaries. By employing force-directed graphing techniques, the diagram dynamically adjusts to illustrate how different concepts within the articles relate to one another, facilitating a clearer understanding of the structural and conceptual linkages that exist across the content. The use of GoJS enables the creation of a highly interactive and visually engaging tool, where users can explore the underlying network of knowledge by interacting with the diagram, thereby providing an insightful representation of the articles' configurational relationships. This method not only enhances the accessibility of the data but also supports the exploration of complex interconnections within the knowledge domain (Figure 4.25).

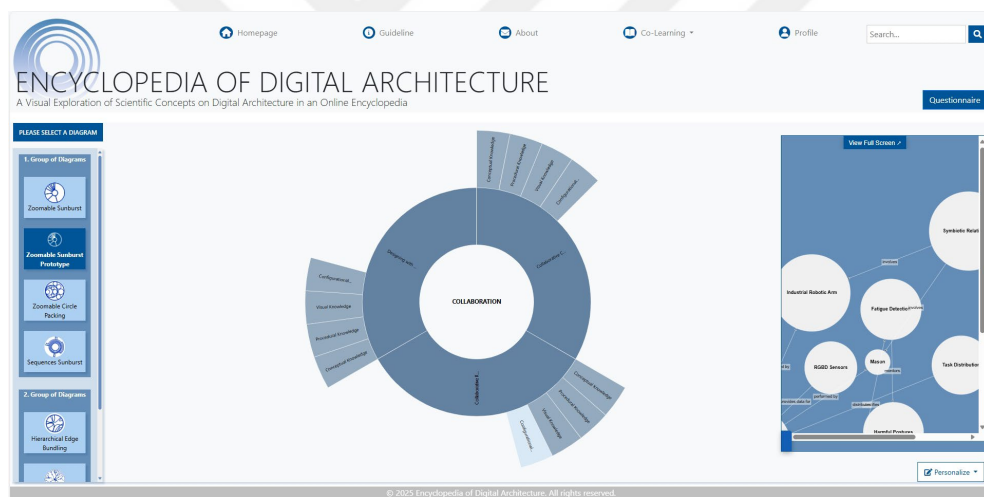


Figure 4.25 : The “Configurational Knowledge” created via GoJS.

4.4.3 Other elements of the web page

The website contains a navigation bar for containing further information. The navigation bar includes a Homepage, an About, a Contact, a Co-Learning, a Profile and a Questionnaire section (Figure 4.26).



Figure 4.26 : The navigation bar of the web page.

The guideline for the web page has been developed to serve a dual purpose: first, to support the implementation of the questionnaire, and second, to assist users in understanding how to effectively navigate and utilize the web interface. The guideline is available in both Turkish and English to ensure accessibility for a broader audience and to facilitate consistent user engagement across different language backgrounds (4.27).

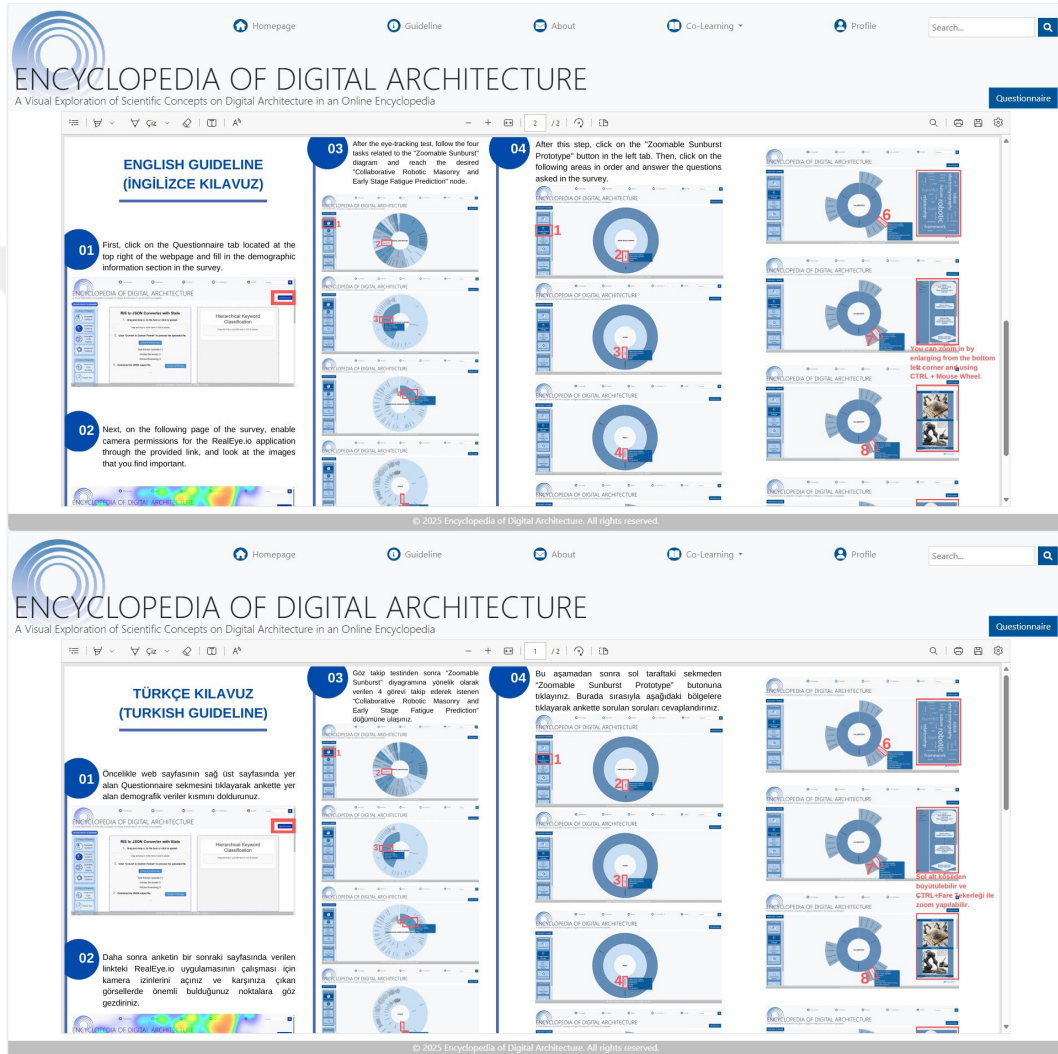


Figure 4.27 : The guideline on the usage of the web page and the questionnaire (in Turkish and English).

The “About” page contains information about the webpage and published articles derived from this thesis as in the Figure 4.28.

The following five visualizations represent additional components of the web interface, each serving a distinct purpose in the representation of knowledge. While all diagrams are derived from the same hierarchical dataset, they differ in structure,

interactivity and the type of cognitive insight they promote. Each visualization offers a unique perspective; some emphasize structural clarity, others prioritize relational depth or narrative sequence, thereby enriching the user's understanding by presenting the same data through varied representational strategies. This multiplicity supports diverse learning styles and enhances the comprehensibility of complex information in the field of digital architecture.

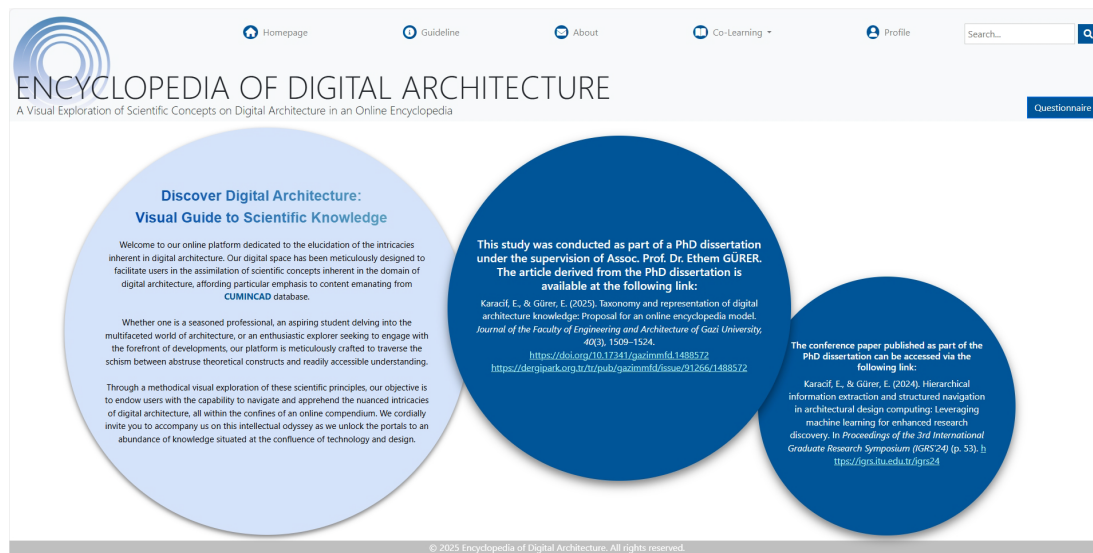


Figure 4.28 : About page detailing the website and the published articles.

The Zoomable Circle Packing diagram, developed using D3.js, offers a hierarchical data visualization method that represents nested relationships through enclosed circles of varying sizes. Unlike the Zoomable Sunburst diagram, which displays hierarchical data in a radial, segmented ring layout, the Circle Packing diagram emphasizes containment and relative size, providing a more intuitive sense of the volume or quantity associated with each node. This visual model allows users to explore data layers by zooming into nested circles, facilitating an understanding of hierarchical depth and scale within clustered data. It is particularly useful for comparing the proportional importance or frequency of categories and subcategories in a compact, non-angular format. By engaging with this visualization, users gain a spatial and quantitative comprehension of how digital architecture knowledge is structured across different thematic levels (Figure 4.29).

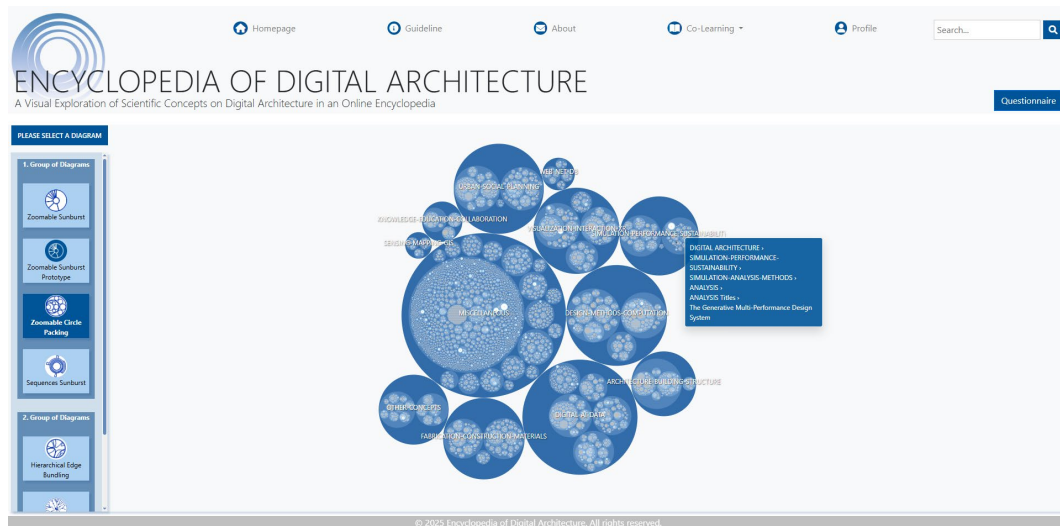


Figure 4.29 : Zoomable Circle Packing diagram created with D3.js for data visualization.

Due to time constraints, the digital architecture knowledge was not fully processed within the following four diagrams. Nevertheless, each was examined with the conceptual framework of digital architecture in mind, despite being populated with alternative content. These diagrams were explored primarily to demonstrate their potential and to assess their capacity to represent complex knowledge structures. As such, they serve as indicative models that offer informative representations into how digital architecture knowledge might be visualized through diverse hierarchical formats.

The Sequences Sunburst diagram, created using D3.js, provides a circular, radial visualization specifically designed to illustrate hierarchical sequences or step-by-step processes within a dataset. Unlike the Zoomable Sunburst or Circle Packing diagrams, which emphasize categorical relationships or proportional containment, the Sequences Sunburst focuses on the linear progression and order of events or decisions across different hierarchical levels. Each arc in the diagram represents a sequential step, and its position and depth indicate the stage and context within the broader structure. This diagram is particularly effective in conveying how digital architecture knowledge evolves procedurally, revealing repeated pathways, branching scenarios, and hierarchical decision trees. By interacting with the diagram, users can trace recurring patterns, offering insights into the knowledge classification and application in the field of digital architecture (Figure 4.30).

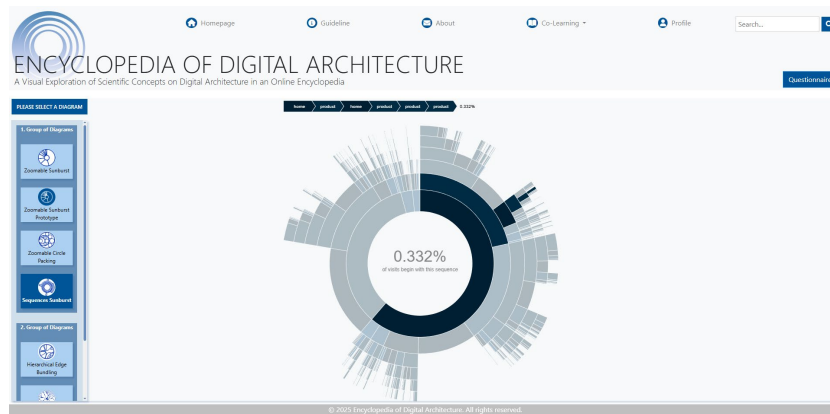


Figure 4.30 : Sequences Sunburst diagram created with D3.js for data visualization.

The Hierarchical Edge Bundling diagram, created with D3.js, is a radial layout specifically designed to visualize complex relationships and interconnections within hierarchical data structures. Unlike the Zoomable Sunburst, which emphasizes hierarchical nesting, or the Sequences Sunburst, which highlights step-by-step procedural flows, the Hierarchical Edge Bundling diagram excels at revealing cross-hierarchical relationships by drawing curved edges between related nodes. These bundled edges reduce visual clutter and highlight patterns of connectivity across different branches of the hierarchy. In the context of digital architecture knowledge, this diagram allows users to explore indirect or latent links between concepts, processes or themes that may not be immediately adjacent in the taxonomy. By visually mapping these associations, the diagram supports a deeper understanding of interdisciplinary overlaps and hidden structures within the knowledge domain, enriching the configurational perspective of the classification framework (Figure 4.31).

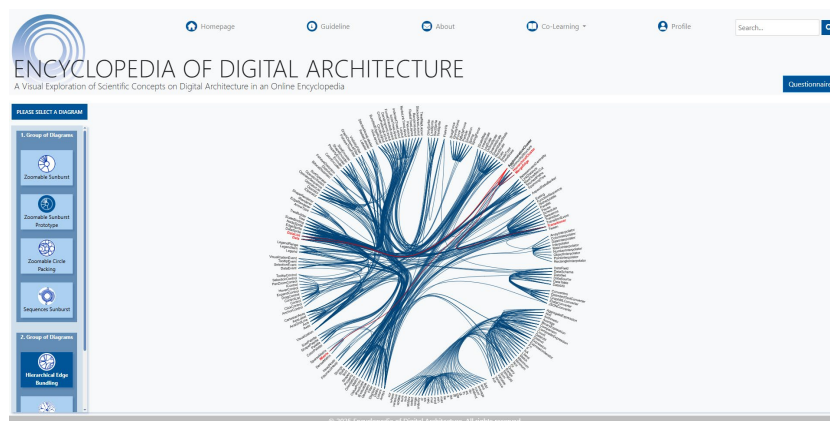


Figure 4.31 : Hierarchical Edge Bundling diagram created with D3.js for data visualization.

The Radial Tree diagram created with D3.js is a circular adaptation of the traditional hierarchical tree structure, designed to visualize complex relationships in a compact and aesthetically organized manner. Unlike linear tree layouts or sunburst diagrams, the Radial Tree distributes nodes evenly around a central point, allowing for a balanced and space-efficient display of hierarchical data. In the context of digital architecture knowledge, this diagram enables users to explore thematic relationships radiating from a central domain, making it easier to perceive both macro-level structures and micro-level subcategories. Its circular configuration enhances readability when dealing with large datasets by minimizing overlapping labels and facilitating interactive navigation. This type of visualization supports a comprehensive understanding of how various knowledge clusters interrelate while offering an engaging visual form that can improve user interaction and cognitive mapping (Figure 4.32).

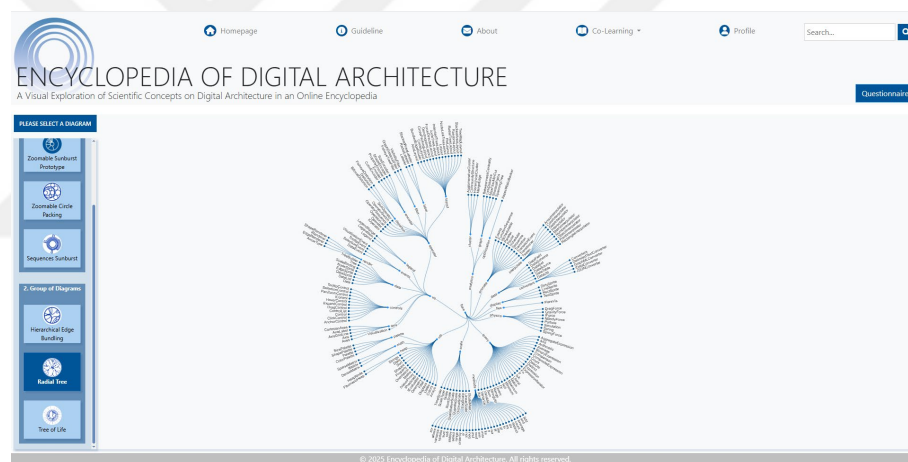


Figure 4.32 : Radial Tree diagram created with D3.js for data visualization.

The Tree of Life diagram created with D3.js is a hierarchical tree layout that visually represents the branching relationships among classified knowledge domains in a clear, top-down or radial structure. Unlike other nested circular or sequential representations, the Tree of Life diagram highlights genealogical or evolutionary relationships between concepts, much like a phylogenetic tree. In the context of digital architecture knowledge, this visualization enables users to trace the lineage and divergence of thematic clusters, observing how broader knowledge categories give rise to more specialized subfields over time. Its clear branching structure makes it especially effective for illustrating parent-child relationships and fostering an intuitive understanding of how different areas of knowledge are interconnected. By

navigating this diagram, users can gain insights into the development and expansion of knowledge in a manner that supports conceptual clarity and a structured overview of the field (Figure 4.33).

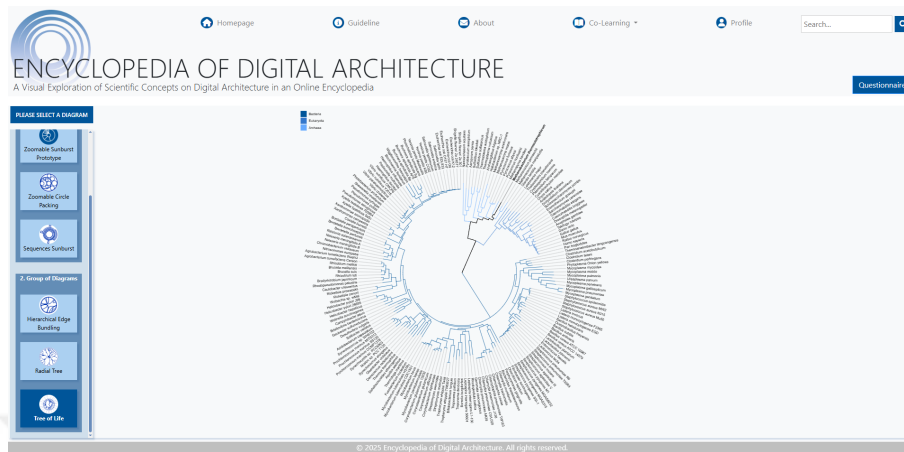


Figure 4.33 : Tree of Life diagram created with D3.js for data visualization.

The testing phase utilizing the full dataset has demonstrated the effectiveness of the developed prototype in facilitating hierarchical visualization and text classification within the domain of digital architecture. The diverse set of D3.js-based diagrams, including Zoomable Sunburst, Zoomable Circle Packing, Sequence Sunburst, Hierarchical Edge Bundling, Tree of Life, and Radial Tree, each offer distinct perspectives on the structural and relational aspects of the classified data. While these visualizations vary in their representation styles and levels of detail, collectively they provide a multifaceted understanding of complex knowledge hierarchies. The prototype's interactive features allow users to explore these layers dynamically, fostering deeper insight into thematic clusters and subclusters. Despite limitations related to time constraints that precluded comprehensive processing of all digital architecture content, the visualization tools successfully illustrate how different diagrammatic approaches can complement one another to enhance the interpretability and accessibility of large-scale text-mined corpora. These findings underscore the potential of integrated hierarchical visualizations as valuable instruments for organizing and communicating domain-specific scientific knowledge.

5. EVALUATION AND RESULTS ON TEACHING DESIGN OF DIGITAL ARCHITECTURE KNOWLEDGE: USER FEEDBACK ANALYSIS

This stage of the study involves a structured, iterative user testing process aimed at evaluating and enhancing the online encyclopedia. To systematically analyze the results, a quasi-experimental design was employed where participant characteristics were treated as independent variables and their performance and perceptions were measured as dependent variables. The primary independent variables included the participants' educational level (e.g., Master's, PhD Student, PhD Graduate) and professional domain (e.g., Architecture, Engineering, Social Sciences). The key dependent variables were: (1) Perceived Usability, measured quantitatively using the System Usability Scale (SUS) score; (2) Learning Performance, measured by the accuracy of responses to the mini-quiz questions for each of the four knowledge layers; and (3) Qualitative Understanding, derived from the thematic analysis of open-ended responses concerning the conceptual, procedural, and configurational tasks ; and (4) Visual Attention Patterns, captured through eye-tracking data, which provided insights into users' focus areas, navigation behavior, and interaction with the visualized knowledge structures.

Each iteration included participants interacting with the current version of the interface, after which comprehensive feedback was obtained via surveys, interviews, and eye-tracking analysis. Based on this feedback, the interface was revised and presented to the subsequent group of users. This cycle was repeated three times to ensure continuous refinement and validation of the design.

The evaluation process specifically examined how participants navigated through the visualization diagrams, including their interaction patterns, intuitiveness of use, and any usability challenges encountered. Insights derived from these observations guided targeted enhancements aimed at improving the user experience and accessibility of the digital platform.

Before beginning their exploration, participants received a brief orientation outlining the purpose of the study and the core functionalities of the online encyclopedia. This

introductory session served as a foundation for meaningful engagement, enabling participants to navigate the system with a clear understanding of its features. During the interaction, participants were encouraged to articulate their impressions, reflect on their user experience, and provide open-ended feedback. Key areas of inquiry included the ease of information retrieval, the effectiveness of visual representations in conveying conceptual interrelationships, and overall perceptions of interface interactivity and usability.

A mixed-method approach was adopted for data collection, combining qualitative input, such as user comments and interview insights, with quantitative data, including standardized satisfaction scores based on the System Usability Scale (SUS). These data streams were synthesized to construct a holistic understanding of the participants' experiences and to inform iterative design decisions.

Additionally, the study employed an iterative testing framework, enabling continuous refinements across testing rounds. This rigorous methodological approach ensured that improvements were grounded in empirical user data, resulting in a nuanced and comprehensive evaluation process. For a detailed list of survey questions utilized in the study, refer to Appendix D.

To further enrich the analysis of user interaction, the RealEye.io platform was integrated into the evaluation process. RealEye is a browser-based tool that leverages artificial intelligence and deep neural networks to perform real-time eye tracking using standard webcams without the need for additional software (RealEye, 2024). Participants were asked to complete a series of tasks while their eye movements were tracked and analyzed. The resulting gaze data was visualized as heatmaps, indicating the areas of the interface that captured the most visual attention.

User testing was carried out with three successive groups. The first group included five participants, followed by a second group of ten, and a third group of five. Each cohort engaged in the full evaluation protocol, combining semi-structured interviews, a mini AI-quiz, user surveys and eye-tracking.

The demographic profile of the participants reflects a diverse but academically inclined cohort. The majority of participants (15 out of 20) fell within the 25–34 age range, while the remaining participants were in the 35–44 age group. In terms of educational background, the participants primarily held advanced degrees: 11 were

doctoral students, 5 held master's degrees, 2 were master's students, and 2 were bachelor's degree holders. Professionally, the group predominantly consisted of individuals from architecture-related fields, including architects, research assistants and academics, along with a smaller representation of professionals from urban planning, sociology, and engineering disciplines. This composition ensured a participant pool with substantial academic and professional engagement in the built environment and digital design disciplines.

Prior to the evaluation, participants were first provided with general guidance on the interface via a dedicated guidelines section embedded within the website. Following this orientation, users were directed to the "Zoomable Sunburst Prototype" section accessible from the left-hand menu. From there, they were guided step-by-step to access the sample article titled "Collaborative Robotic Masonry and Early Stage Fatigue Prediction". Participants were then instructed to sequentially explore each of the four main knowledge categories (conceptual, procedural, visual, and configurational) by actively clicking on their respective nodes within the visualization. At each stage, they were prompted to respond to open-ended questions designed to assess their comprehension and interpretative engagement with the knowledge structure.

In this context, during the first phase of the evaluation, semi-structured interviews were conducted, and the responses to the question regarding Conceptual Knowledge, which asked users to identify and recall at least two primary categories and three of their subcategories based on the sample article, demonstrated an active cognitive engagement with the subject matter. The most frequently recalled core categories included "masonry, robotics and framework" while the most commonly cited sub-layers involved "tool, posture, relationship, fatigue and harmful." The recurrence of these terms across responses suggests that participants were able to meaningfully interpret and internalize the core conceptual structures embedded in the selected article. These structures predominantly revolve around the themes of collaborative robotic masonry and early-stage fatigue prediction, reflecting participants' understanding of key issues such as human-robot interaction, ergonomics, and the implications of automation in construction workflows. The uniformity in terminology across different participants' responses further underscores the

effectiveness of the interface in conveying conceptual hierarchies and supporting structured knowledge acquisition through the interactive prototype.

Following their interaction with the Conceptual Knowledge layer, participants proceeded to explore the Procedural Knowledge component within the Zoomable Sunburst Prototype interface. In this stage, users were presented with a process diagram illustrating the operational sequence of the selected article. Participants were asked to interpret the flow diagram, summarize it in a maximum of five key stages, and infer the overall aim of the study. In addition, they were encouraged to reflect on whether a visual representation of the process emerged in their mind prior to transitioning to the Visual Knowledge layer, and if so, to describe or draw it.

The responses revealed a consistent understanding of the procedural structure underpinning the study. Participants frequently articulated a sequential process that involved monitoring the construction worker's posture, analyzing real-time data, identifying harmful ergonomic patterns or fatigue indicators, issuing corrective suggestions or robotic interventions, and supporting the human worker in optimizing movement or reducing strain. Across responses, a shared interpretation of the study's aim emerged: the development of a collaborative system wherein robotic tools assist human masons by improving posture, reducing fatigue, and increasing long-term work efficiency.

Participants also described vivid mental images prompted by the procedural flow. Many envisioned a construction worker interacting closely with a robotic arm or wearable device, where the robot observed and analyzed bodily movements and provided ergonomic feedback or support. This image was often framed in terms of a symbiotic or assistive relationship, such as a robot stabilizing a worker's spine or guiding their arm to prevent injury. These mental models indicate that the procedural diagram successfully activated anticipatory visual thinking, reinforcing the link between procedural understanding and visual cognition.

Furthermore, a number of participants expanded the scope of their interpretations by drawing connections to broader architectural or philosophical themes. For instance, some reflected on the increasing integration of robotics into skilled labor, likening the robot-human collaboration to a "cyborg" condition that redefines the role of

craftsmanship in the construction process. Others interpreted the system as a form of extended cognition or embodied intelligence in the architectural production pipeline.

Overall, the responses suggest that the procedural representation in the interface was effective in facilitating the comprehension of sequential workflows while also stimulating speculative visualizations and conceptual inferences. This indicates the potential of structured procedural knowledge to serve as a cognitive bridge between abstract conceptual domains and concrete visual understanding in digitally mediated learning environments.

The Visual Knowledge layer significantly contributed to the participants' understanding of the article's overall narrative by offering intuitive and perceptible representations of abstract and procedural content. Responses reveal that visual materials, particularly technical drawings, pictograms, and sketches, were not only instrumental in reinforcing earlier layers of knowledge but also served as key communicative tools in making complex concepts more accessible.

Among the four primary visual formats presented, technical drawings and pictograms were frequently regarded as the most meaningful. Participants emphasized that technical drawings effectively illustrated human robot collaboration by showcasing spatial relations and human-scale references. The inclusion of human figures performing specific tasks enabled a direct visual correlation with procedural elements described earlier in the article. In contrast, pictograms were appreciated for their aesthetic clarity, use of color and capacity to abstractly represent key components of the hybrid construction process. Their symbolic nature appeared to facilitate rapid comprehension of roles, processes, and spatial relationships.

Additionally, some participants found the sketch representations particularly impactful. These visuals were seen as more expressive and human-centered, blending conceptual imagination with technical intent. The use of sketching as a medium was perceived to implicitly preserve the presence of the human author and emphasize the imaginative dimension of architecture and robotics.

The majority of participants affirmed that the visual materials were coherent with the information presented in the Conceptual and Procedural layers. In particular, the visuals were said to align closely with the narrative concerning human-robot interaction, task delegation, and construction scenarios, thereby reinforcing key ideas

such as symbiotic collaboration, posture monitoring and supportive automation. Visuals functioned not only as supplements but as essential interpretive bridges, allowing participants to synthesize information from previous layers through a more embodied lens. A few participants mentioned that some visuals did not clearly reflect the content of the text or were only partially relevant. This suggests that including more detailed explanations or context-specific labels in the visuals could improve their clarity and alignment with the written content.

In conclusion, the Visual Knowledge layer effectively augmented the Conceptual and Procedural layers by transforming abstract concepts and procedural information into intelligible, symbolically rich, and human-centric visual representations. The inclusion of diverse visual modalities enhanced interpretive depth, facilitated cognitive synthesis and accentuated the hybridized character of digital construction processes. The coherence between these visualizations and the preceding knowledge layers, along with their capacity to articulate collaborative human-machine dynamics, highlights their pedagogical significance within the epistemological framework of the article.

The posed question aimed to explore the Configurational Knowledge layer by analyzing a network mapping and inviting associative expansion through the addition of three key ideas linked to the existing nodes. This approach reflects the layer's emphasis on relational understanding, systemic arrangement and the interconnectivity of concepts within a knowledge structure.

The responses provided demonstrate a multidisciplinary engagement with the configurational dimension of the subject matter. They extend beyond mere thematic labeling, offering nuanced insights into how disparate concepts, ranging from ergonomics and harmful postures to machine learning, real-time data analysis, and augmented reality, can be interlinked within an integrative knowledge framework. This reflects a mature grasp of configurational knowledge as a dynamic, networked system rather than a static collection of isolated facts.

Notably, the contributions emphasize practical and technological integration, such as proposing machine learning models for fatigue prediction, mobile alert systems for posture correction and data visualization tools for tracking individual performance metrics. These suggestions underscore the procedural and applied implications

embedded within configurational thinking, where knowledge is not only mapped but actively operationalized to solve real-world problems. Participants also suggested extensions to the current framework, including algorithm optimization for faster response, inclusion of health monitoring through emotional and pain tracking, and expansion of robot-supported urban design and waste management tasks. These reflections point to a forward-looking perspective where configurational knowledge is not static but evolves through iterative processes and interdisciplinary collaboration.

Furthermore, the responses demonstrate a clear awareness of the importance of contextual applicability. For instance, the suggestion to move experiments from laboratory settings into real-world environments aligns with current expectations for enhancing practical relevance and applicability.

Subsequent to the semi-structured interviews, mini AI quizzes were administered. The mini AI quizzes designed to assess participants' understanding of conceptual, procedural, visual, and configurational knowledge revealed notable variations in performance across these categories. All 20 participants correctly answered the questions related to Conceptual and Visual Knowledge, indicating a strong and consistent grasp of fundamental concepts and visual representations within the studied domain. This suggests that the instructional materials and interface effectively conveyed both the theoretical frameworks and the corresponding visualizations, enabling comprehensive knowledge acquisition in these areas.

In contrast, Procedural Knowledge questions showed a marginally lower accuracy, with a single participant providing an incorrect response. This slight decline may reflect the increased cognitive complexity involved in understanding sequential workflows and operational processes compared to static concepts or visual structures. It also highlights the importance of reinforcing procedural sequences to ensure a robust and uniform comprehension among users.

Configurational Knowledge, meanwhile, yielded the highest number of errors, with four incorrect answers recorded. This relative difficulty underscores the layered and systemic nature of configurational understanding, which involves integrating multiple components such as human factors, technological interoperability and contextual adaptability. The errors suggest that participants may require additional

support or clearer guidance to fully internalize the dynamic and interconnected aspects inherent in configurational knowledge. Overall, these results emphasize the need for tailored pedagogical strategies that address the distinct cognitive demands of each knowledge type to optimize learning outcomes in digitally mediated architectural contexts.

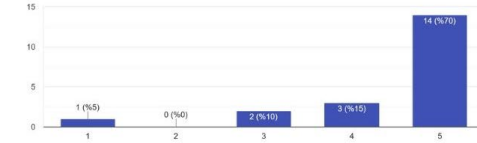
After completing the mini AI quizzes, participants were asked to complete the SUS (System Usability Scale) survey. The System Usability Scale (SUS) scores were calculated following the standard procedure: for positively worded items (1, 3, 5, 7, 9), the score was computed by subtracting 1 from the user rating, whereas for negatively worded items (2, 4, 6, 8, 10), the score was derived by subtracting the user rating from 5 (Figure 5.1). These adjusted scores were then summed and averaged per participant, and subsequently multiplied by 2.5 to convert the results into a 0–100 scale.

Based on the aggregated responses from 20 participants, the average SUS score falls within the range of approximately 75 to 80. This range indicates a good level of usability, as scores above 68 are generally considered above average, and scores between 75 and 80 reflect a system that users find highly usable, easy to learn and efficient in operation.

In addition to the average SUS scores, the standard deviations were examined to assess the consistency of user experiences across participants. The relatively low standard deviations observed (generally ranging between 0.8 and 1.5 on the 5-point Likert scale items) indicate that the responses were fairly homogeneous, reflecting a consistent perception of the system's usability among users. This consistency reinforces the reliability of the SUS findings, suggesting that the majority of users shared similar positive evaluations of the system's ease of use, learnability, and efficiency. However, slightly higher variability in responses to negatively phrased items related to technical support and perceived complexity implies that individual differences may exist in users' comfort levels or prior experience with similar systems, highlighting potential areas for targeted user assistance or interface refinement.

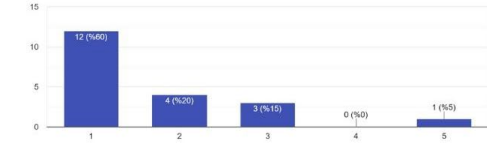
1. I think that I would like to use this system frequently.

1. Bu sistemi daha sık kullanmak istiyorum.
20 yanıt



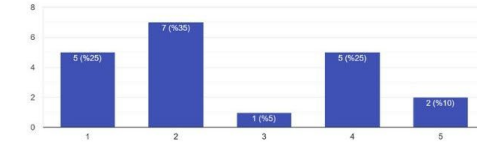
6. I thought there was too much inconsistency in this system.

6. Bence sistemde çok fazla düzensizlik var.
20 yanıt



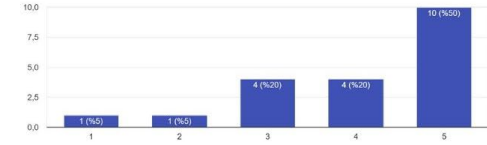
2. I found the system unnecessarily complex.

2. Bu sistemin olması gerekenden daha karmaşık olduğunu düşünüyorum.
20 yanıt



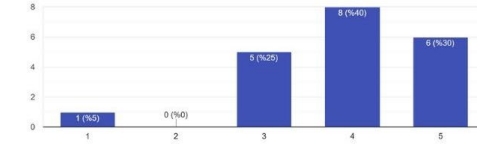
7. I would imagine that most people would learn to use this system very quickly.

7. Bence çoğu insan bu sistemi hızlı bir şekilde öğrenebilir.
20 yanıt



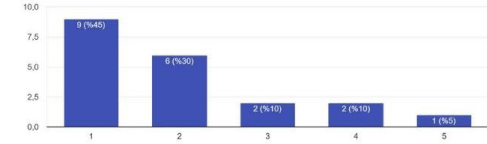
3. I thought the system was easy to use.

3. Sistemin basit ve kullanımı kolay olduğunu düşünüyorum.
20 yanıt



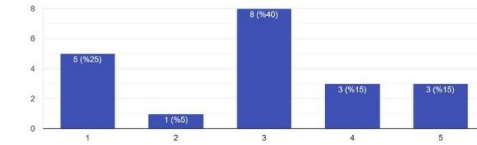
8. I found the system very cumbersome to use.

8. Bu sistemin zaman alıcı olduğunu düşünüyorum.
20 yanıt



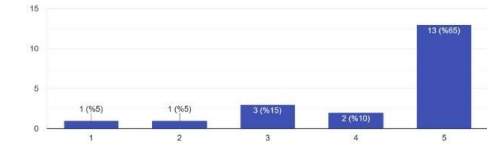
4. I think that I would need the support of a technical person to be able to use this system.

4. Bu sistemi kullanmak için teknik desteğe ihtiyacım var.
20 yanıt



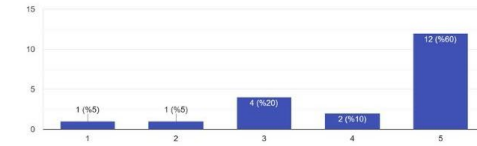
9. I felt very confident using the system.

9. Bu sistemi kullanırken kendimi güvende hissediyorum.
20 yanıt



5. I found the various functions in this system were well integrated.

5. Sistemin sorunsuz çalıştığını ve iyi entegre edildiğini görüyorum.
20 yanıt



10. I needed to learn a lot of things before I could get going with this system.

10. Bu sistemi kullanmaya başlamadan önce öğrenmem gereken çok şey olduğunu düşünüyorum.
20 yanıt

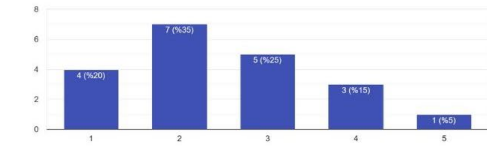


Figure 5.1 : Graphical representation of SUS results.

The data shows a strong positive consensus on core usability and learnability aspects. For instance, Item 3 ("I thought the system was easy to use") and Item 7 ("I felt that most people would learn to use this system very quickly") consistently received high scores (mostly 4s and 5s), demonstrating the platform's intuitive design. Similarly, Item 1 ("I would like to use this system frequently") and Item 9 ("I felt very confident using the system") also scored highly, suggesting users not only found the system effective but also felt comfortable and empowered while using it.

Conversely, items designed to measure negative attributes, such as Item 2 ("I found the system unnecessarily complex") and Item 6 ("I thought there was too much inconsistency in this system"), received consistently low scores (mostly 1s and 2s). This strongly suggests that the hierarchical structure and visual interface were perceived as logical and coherent, not confusing.

Interestingly, the highest variability in responses (i.e., higher standard deviation) was observed for Item 4 ("I think that I would need the support of a technical person to use this system") and Item 10 ("I felt I would need to learn a lot of things before I could get going with this system"). While many users felt confident, a notable minority expressed a need for guidance. This points not to a flaw in the system's design, but to the inherent complexity of the subject matter, digital architecture itself. It suggests that while the interface is usable, the conceptual depth of the content might prompt some users, particularly those less familiar with the domain, to seek further clarification. This distinction is crucial as users did not find the tool difficult but acknowledged the richness of the knowledge it presents. This finding reinforces the platform's success as an effective environment for exploring complex information, while also highlighting potential areas for incorporating more embedded guidance or tutorials for novice users.

In summary, the SUS results demonstrate that users perceive the system as well-integrated and reliable, with generally smooth operation and a friendly user interface. While minor issues related to learning support and time investment were noted, the findings suggest that the system has achieved a level of usability that supports effective adoption and practical application in its intended context.

Following the analysis of SUS results, the eye-tracking findings were examined. The generated heatmaps revealed clear patterns of user attention. In the first group, three out of five participants focused predominantly on two plugin-related content sections that were deliberately designed to attract visual interest. The other two users primarily concentrated on the upper-left quadrant of the page, aligning with common web-navigation behavior.

Across all groups, the most visually salient elements were consistently the navigation bar at the top of the interface, followed by the left-side vertical button panel, and finally, the plugin content sections. These findings provide critical insight into the

website's visual hierarchy and offer an empirical basis for optimizing design elements in alignment with natural gaze behaviors (Figure 5.2). Additional heatmaps generated for other key interface pages during the first testing round are included in Appendices A1 through A5, in order to maintain the readability and flow of the main text.

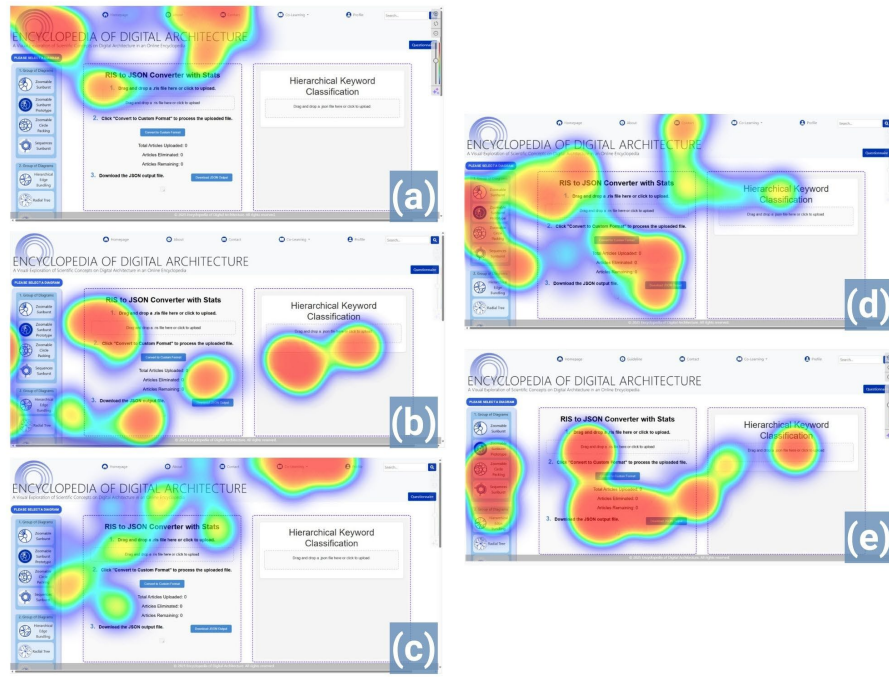


Figure 5. 2 : Heatmap results from group 1 on the homepage (RealEye, 2024).

Analysis of eye-tracking data from the second group revealed three distinct user archetypes based on visual engagement patterns. The first cohort (participants "c," "e," and "i") exhibited a top-down scanning behavior, concentrating their gaze on the upper portion of the webpage, which suggests a prioritization of header information and top-level navigation. In contrast, the second cohort (participants "a," "b," "d," and "f") demonstrated a comprehensive exploratory pattern, methodically scanning nearly all significant interface elements. This indicates a more thorough and systematic approach to information gathering.

A third, particularly insightful pattern emerged from participants "g," "h," and "j," who displayed a content-centric focus. While this group largely bypassed the left-side navigation panel, their attention was intensely concentrated on the core interactive plug-in sections. This behavior suggests that peripheral navigation may be secondary for users who have identified the primary content area, underscoring the central content's primacy in anchoring user focus.

Across all participants in this group, a general trend was observed where gaze distribution originated from the webpage's center and radiated outward, consistent with natural reading and scanning behaviors. These differentiated findings provide nuanced insights into user navigation strategies and offer an empirical basis for design refinements aimed at improving discoverability and optimizing the interface layout (Figure 5.3). To maintain narrative flow, heatmaps for other key pages from the Group 2 analysis are provided in Appendices B1 through B5.

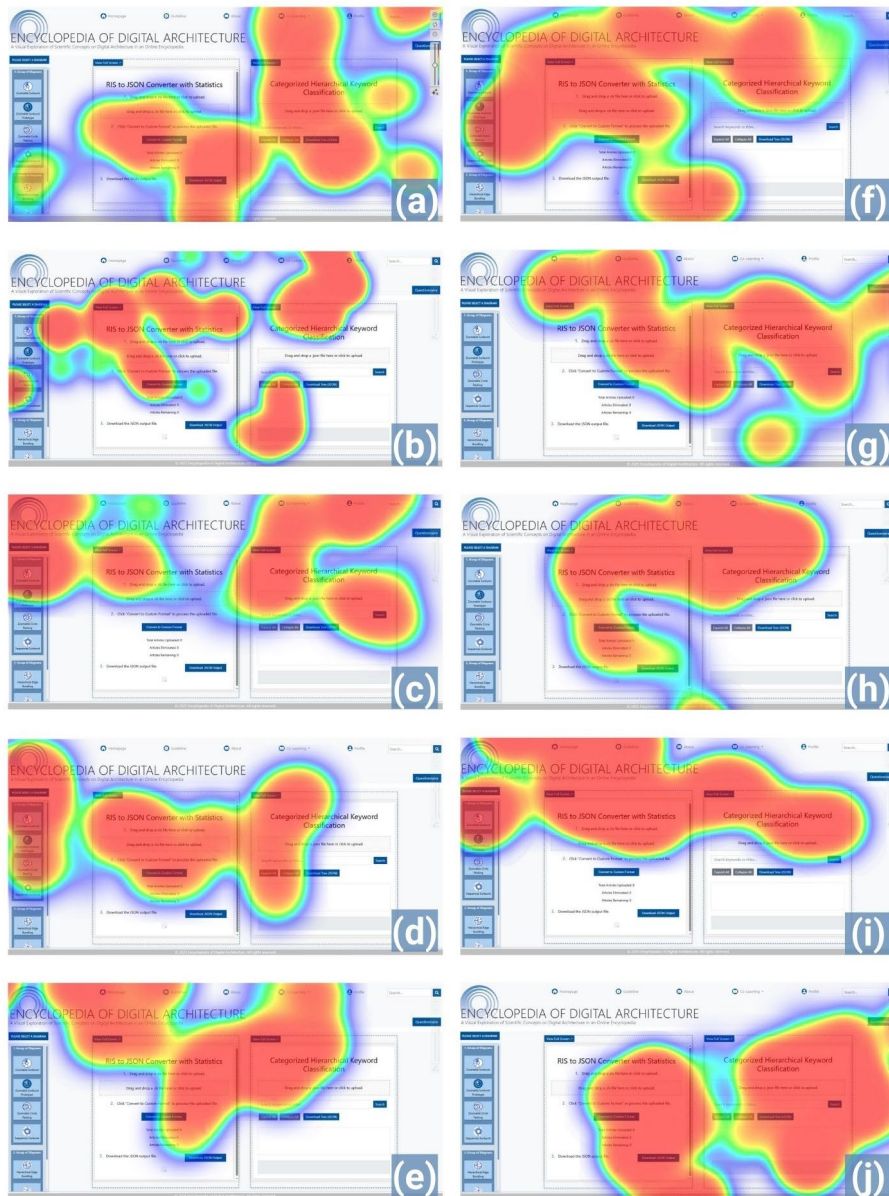


Figure 5.3 : Heatmap results from group 2 on the homepage (RealEye, 2025).

Eye-tracking analysis of the third group revealed heterogeneous gaze patterns, which can be categorized into distinct behavioral cohorts. One cohort, comprising participants "c," "d," and "e," adopted a comprehensive scanning strategy, with their

visual attention distributed across the entirety of the webpage. This holistic pattern suggests a thorough exploratory approach aimed at assimilating all available information before focusing on specific tasks.

In contrast, other participants demonstrated more targeted engagement. Participant "a" concentrated visual attention on the upper-central region, indicating a focus on primary header information or a prominent content anchor. Similarly, participant "b" exhibited highly selective attention, allocating sustained focus to the central plug-in menus while giving only cursory glances to the peripheral left-hand navigation. This behavior signifies a task-oriented approach, where core interactive elements are prioritized over secondary navigational aids.

Despite these individual variations, a unifying trend emerged where the central region of the page consistently commanded the highest intensity of visual focus across all participants. This finding unequivocally underscores the primacy of the central content area in capturing and guiding user interaction. These observations provide compelling empirical support for design strategies that place critical information and interactive components within the central visual field to maximize user engagement and streamline task completion (Figure 5.4). To maintain narrative continuity, supplementary heatmaps for this group are provided in Appendices C1 through C5.



Figure 5. 4 : Heatmap results from group 3 on the homepage (RealEye, 2025).

Finally, user feedback was carefully reviewed and evaluated to identify potential areas for improvement. Based on the responses, several enhancements were implemented to refine the system's usability and user experience. For instance, users suggested increasing the window size for knowledge categories to facilitate better readability. Additionally, incorporating more personalization options within the system interface was recommended to further enhance the user experience. Many participants appreciated the system's ability to filter information efficiently, thereby reducing research time.

Regarding visual design, some users requested clearer and updated images, noting that certain diagrams and supplementary graphics, especially those on the right side of the page, could be presented in larger, more accessible frames. While the system initially appeared complex and overwhelming to some participants, they acknowledged that familiarity increased with time. The comprehensive background work was recognized, though clearer guidelines and more intuitive visuals were deemed necessary. Furthermore, several users emphasized the importance of mobile accessibility, suggesting the development of mobile-friendly versions or dedicated applications to enhance practical usability.

In addition, some users expressed that text sizes could be increased for improved readability. Overall, the constructive feedback has informed ongoing development efforts to optimize both the interface and interaction design, aiming to create a more intuitive, accessible and efficient system for all users.

6. DISCUSSIONS AND LIMITATIONS

6.1 Evaluating the Research Objectives

6.1.1 The potential of hierarchical structures for exploratory learning

One of the central research questions of this thesis was whether hierarchical structures, as opposed to linear text formats, could facilitate deeper and more meaningful learning in the context of digital architecture. The findings of the study affirm this hypothesis. Hierarchical structures, when combined with interactive visualizations, support a form of exploratory learning that enables users to perceive the relationships between different knowledge components rather than processing them as isolated facts.

The categorization of knowledge into four layers (conceptual, procedural, visual, and configurational) allowed users to approach digital architecture knowledge from multiple perspectives. This structure encouraged associative thinking and facilitated cognitive mapping, which are essential for understanding complex and interconnected concepts typical of digital design environments. The ability to navigate back and forth between related categories and zoom into subtopics enhanced engagement and retention, as also evidenced by user feedback and eye-tracking data.

Furthermore, hierarchical models offered users a sense of orientation within the knowledge domain. Unlike conventional text-based learning, where information is consumed sequentially and often forgotten once read, the layered and networked structure of the proposed encyclopedia enabled users to construct their own pathways through knowledge. This aligns with constructivist learning theories, which emphasize learner agency, contextual understanding, and active meaning-making.

The eye-tracking results indicated that users spent more time on sections where visual hierarchy was evident and intuitively navigable, such as the sunburst diagram and the categorized article views. This suggests that users were not only passively reading but actively exploring and evaluating the material. In conclusion, the

hierarchical structuring of digital architecture knowledge proved to be a powerful pedagogical strategy, reinforcing the value of non-linear, visually navigable learning environments.

6.1.2 The role of systematic classification and relational visualization

The second research question explored the potential of systematic classification and relational visualization in enhancing the accessibility and usability of digital architecture knowledge. The empirical findings substantiate that this combined approach significantly contributes to both user experience and cognitive engagement. The system achieved a high average score on the System Usability Scale (SUS), ranging between 75 and 80, indicating a strong overall usability. Beyond quantitative measures, qualitative data underscored the pivotal role of visual classification in facilitating effective knowledge navigation.

Specifically, the force-directed graph employed for configurational knowledge enabled users to uncover latent, non-obvious connections between diverse research themes. This process of visual discovery transforms the act of learning from a passive reception of information to an active inquiry and exploration. By making the relational structure of the knowledge domain explicit, visualization supports deeper understanding and meaningful cognitive processing.

The study demonstrates that systematic classification, when integrated with interactive visualization techniques, transcends mere information organization; it fundamentally enhances knowledge accessibility and cognitive utility. Revealing the inherent relational patterns within the knowledge domain enables users to intuitively grasp complex interdependencies, thus fostering more profound insights and facilitating advanced problem-solving. Consequently, this integrative approach offers a novel paradigm for managing digital architecture knowledge, promoting more effective dissemination and exploratory learning within online scientific encyclopedias.

6.2 System Limitations and Data Challenges

This research contributes to the field of digital architecture by proposing a layered classification of digital architecture knowledge and visualizing it through a web-based interface. The study demonstrates how the integration of AI-based text mining

and hierarchical visualization techniques can aid in structuring, interpreting, and communicating complex bodies of knowledge. One of the strengths of this study lies in the systematic differentiation between conceptual, procedural, visual, and configurational knowledge, which not only reflects the multidimensional nature of computational thinking but also informs how digital tools can accommodate and enhance such complexity. However, several limitations should be acknowledged, as developing an online visual encyclopedia within the scope of this study represents a multifaceted challenge.

First, the articles selected for classification are exclusively retrieved from a single source, CUMINCAD, a comprehensive repository of scientific publications. Broader databases such as Web of Science or Scopus were excluded from the scope of this thesis due to the extensive volume of data they encompass. Incorporating such databases would pose significant challenges, including the time-intensive process of data collection and the complexity of managing an extensive and intricate web of information. Importantly, CUMINCAD is deemed sufficient for illustrating the relationships and connections within the domain of digital architecture, providing a focused and relevant dataset for the purposes of this study.

Second, the creation of an automated system for knowledge extraction and classification necessitates either the utilization of external APIs or the development of complex code structures from the ground up. While the current system implements custom logic for keyword processing and hierarchical structuring, reliance on certain external APIs for more advanced NLP tasks (e.g., sophisticated summarization or entity recognition beyond the current scope) can present challenges. For instance, APIs like those available through platforms such as OpenAI, Gemini Developer API or Hugging Face, while powerful, may impose rate limits or usage costs that can hinder continuous, large-scale data processing or real-time system functionality. This can disrupt the seamless operation of the system and potentially lead to performance issues if not carefully managed.

Third, the development effort was constrained by a limited familiarity with a broad range of advanced coding tools and frameworks. While the current JavaScript-based implementation is robust for its intended purpose, leveraging more specialized data science libraries or back-end technologies might have offered alternative pathways for optimization or feature expansion. This practical limitation influenced the choice

of client-side processing, which, while offering accessibility, might face performance bottlenecks with extremely large datasets compared to server-side architectures.

Fourth, a current limitation of the system lies in its data-to-visualization pipeline, which is not yet fully automated. At present, the generation of the JSON file that feeds the interactive visualizations involves manual steps, creating a lag between data analysis and its public representation. Future work should focus on creating a seamless, end-to-end workflow where AI-driven analysis automatically generates and updates the JSON data structure. Achieving this level of automation would not only ensure near real-time content updates but also significantly enhance the system's scalability and robustness, a goal that represents the next stage of the study's technical development.

Finally, a notable limitation of the current web interface is its lack of full responsiveness across diverse screen sizes and devices. While designed for desktop use, its optimal rendering and usability on smaller screens, such as tablets or mobile phones, have not been systematically addressed. This impacts the accessibility of the tool for users on a wider range of platforms.

Despite these constraints, significant efforts were directed toward ensuring the website operates with a high degree of automation for keyword classification and visualization to facilitate the seamless interpretation of knowledge. This approach aimed to balance technical feasibility with the objectives of creating an accessible and efficient tool.

6.3 Future Directions and Research Potential

Building upon the current system, several avenues for future research and development present themselves. A primary direction involves expanding the data sources beyond CUMINCAD to include major scholarly databases like Web of Science and Scopus. This would provide a more comprehensive and potentially diverse view of the digital architecture knowledge landscape, though it would necessitate robust strategies for data acquisition, cleaning and integration, as well as potentially more scalable backend processing.

Enhancing the NLP capabilities is another crucial area. Integrating more advanced techniques for semantic analysis, such as topic modeling (e.g., Latent Dirichlet

Allocation) or word embeddings (e.g., Word2Vec, BERT), could uncover latent thematic structures and relationships between keywords and documents that are not apparent through direct keyword matching and predefined categorization alone. This could lead to a more nuanced and data-driven classification system, potentially reducing reliance on manually curated synonym and category maps.

The visualization component offers significant potential for improvement. Future work could explore the integration of more diverse and interactive visualization paradigms, such as dynamic network graphs to illustrate co-occurrence of keywords or collaborations. Implementing advanced filtering and faceting options based on publication year, author or specific sub-themes would also enhance the exploratory capabilities of the interface. Ensuring full responsiveness for mobile and tablet devices is also a key development goal to improve accessibility.

From a user interaction and experience perspective, the system's evolution can be envisioned along a trajectory of increasing intelligence and immersion. A foundational step is to incorporate features for community-driven enrichment, allowing users to provide feedback, suggest new synonyms, or collaboratively annotate relationships, thereby transforming the system into a dynamic, living knowledge base. Building upon this, future development can focus on three advanced capabilities: adaptive learning, interactive dialogue, and immersive exploration. First, adaptive learning could be achieved by implementing artificial intelligence algorithms that personalize content delivery. This would move beyond simply allowing users to save views, instead actively tailoring knowledge pathways and visualizations based on their individual profiles and learning objectives. Second, interactive dialogue could be fostered by integrating sophisticated generative AI chatbots. This would empower users to converse with the knowledge base, pose complex queries in natural language, and receive synthesized textual analyses of articles on demand. Finally, immersive exploration represents the most ambitious trajectory, where the study's scope expands to include virtual and augmented reality. This would allow for the transformation of data visualizations into three-dimensional, explorable "knowledge landscapes," with the potential integration of gamification mechanics to create a more engaging and motivating framework for discovery.

By analyzing data across multiple time periods, the system can effectively track the progression of research themes, the emergence of novel concepts and transformations in the intellectual landscape of the digital architecture field over time, thereby offering meaningful insights for students, researchers and educators alike.

Additionally, enhancing the system's functionality through more effective integration of APIs could facilitate the simultaneous generation of the visual, procedural and configurational knowledge layers. Furthermore, incorporating Natural Language Processing (NLP) APIs is expected to positively contribute to the system's synchronicity, thereby improving the overall efficiency and coherence of the knowledge representation process.



7. CONCLUSIONS

This thesis introduced and validated a novel framework for the classification and visualization of digital architecture knowledge, demonstrating that a structured, multi-layered approach can significantly enhance exploratory learning and digital archiving. By moving beyond conventional, linear modes of information delivery, this study makes a fundamental contribution to the pedagogy of digital architecture. It establishes that a domain's knowledge is not monolithic but is comprised of distinct yet interconnected layers: the conceptual (theories and ideas), the procedural (methods), the visual (representations), and the configurational (systemic relationships and structures). The research established this framework by first leveraging established text mining methods on the CUMINCAD database to create a structural backbone, and then transcending the limitations of initial algorithmic categorization through generative AI to achieve nuanced, context-aware interpretations. Fundamentally, this research operates on the epistemological premise that knowledge is not merely a collection of facts but a structured, relational entity. By visualizing this structure, the study transcends simple information retrieval and enables structural knowing, an understanding of how concepts, methods, and representations interrelate to form a coherent disciplinary framework. This perspective is the core intellectual contribution of the thesis, positioning it not just as a technological development but as an inquiry into the nature of digital architecture knowledge itself.

A key aspect of this framework is the synergistic integration of visualization and textual analysis. While this research confirms that data visualization enhances accessibility by enabling a quick grasp of trends and patterns, it also acknowledges that it may not capture the full richness derived from detailed textual analysis. This acknowledges the inherent trade-off in any knowledge representation, as it requires balancing the holistic and pattern-revealing power of visual abstraction with the nuanced depth of textual narration. The proposed model offers a robust pathway to a

comprehensive assimilation of knowledge by using visual insights to guide textual exploration and vice versa. The diverse visual narratives, from zoomable sunbursts to relational graphs, allow users to interact with data dynamically, fostering a gradual and comprehensible abstraction of scientific information.

This study opens new directions for the design of educational technologies in digital architecture. The proposed model offers a foundation for further exploration of how digital tools can reshape knowledge production, structuring, and sharing in design-oriented disciplines.

This research established a novel framework for structuring, synthesizing, and mapping scientific knowledge in digital architecture. The foundation for this approach was the CUMINCAD database, whose comprehensive article metadata served as the primary resource. Initial analysis relied on established text mining methods, including TF-IDF vectorization and the unsupervised learning algorithm k-Means clustering, to create a primary structural organization of the knowledge domain.

Building upon this structural backbone, the study's primary contribution lies in the subsequent application of generative AI. This advanced step transcended the limitations of initial algorithmic categorization, demonstrating a robust method for achieving nuanced, context-aware interpretations of the large scholarly corpus. Furthermore, the development of custom plug-ins represents a critical step towards realizing a dynamic and living encyclopedia. This feature transforms the system from a closed repository into an extensible platform, empowering the community to ensure its continued relevance through ongoing data contributions and fostering a sustainable model for scholarly knowledge curation.

The originality of this work, therefore, lies not in the individual application of text mining or visualization, but in their synergistic integration toward a specific pedagogical goal. While other tools may classify or visualize data, this research is unique in its development of a multi-layered classification schema (conceptual, procedural, visual, and configurational) that is itself a theoretical proposition about the constituent parts of digital architecture knowledge. This framework provides a new lens through which the field can be understood, analyzed, and taught, moving

the discourse from a focus on discrete digital tools to a more holistic understanding of the computational design thinking process.

The study encompasses an interdisciplinary approach combining digital humanities and architecture, potentially transforming digital architecture education. It has generated a dynamic platform that allows users to explore and understand the expanding structure of digital architecture. Through deeper analysis and visual representations, the study surpasses traditional methods of knowledge transmission by providing concise and instant summaries, enhancing accessibility and comprehensibility. The proposed hierarchical structure enables users to navigate systematically from broad concepts to specific details.

While data visualization enhances accessibility and understanding by enabling quick grasping of trends or patterns, it may not capture the richness and complexity derived from detailed textual analysis. Opting for data visualization is beneficial in scenarios where rapid comprehension and access to summarized visual information are crucial. However, for deeper and nuanced understanding, critical analysis, and comprehensive evaluation of research, reading the full article is necessary. Nonetheless, the focus of this study is on integrating a comprehensive perspective of scientific articles and the concepts utilized within the field of digital architecture.

The diagrams used provide diverse visual narratives, enabling users to access various perspectives on scientific information related to digital architecture according to their preferences. This gradual abstraction of scientific information makes it more comprehensible. Each type of graphic representation has its visual aesthetic, allowing users to experience the details of information through the representation that appeals to them most. Users can interact with datasets from the scientific database of digital architecture using dynamic visuals, zooming in and out, and focusing on specific data.

The relative effectiveness of obtaining information using data visualization versus reading an article in its entirety depends on the reader's goals and the type of information being sought. Although data visualization can provide a succinct and instantaneous synopsis, improving accessibility and understanding, it could not have the richness and complex background that come from an extensive textual analysis.

Data visualization serves as a powerful cognitive tool for efficiently processing complex information and perceiving underlying trends or patterns. Visual representations, such as graphs and charts, enable users to synthesize intricate data linkages far more rapidly than is possible with purely textual or numerical formats. Crucially, this method is not a shortcut for a more general comprehension under time constraints. Rather, it facilitates a fundamental shift from linear, sequential analysis to a holistic, systemic understanding, allowing for the intuitive discovery of relationships that might otherwise remain obscured within the data.

Therefore, the choice is not between data visualization or linear reading, but rather how to best leverage their complementary strengths. The selection of an approach should be guided by the specific analytical goals and pragmatic constraints of the inquiry. While visualization excels at revealing holistic structures and systemic patterns, in-depth textual analysis provides the nuanced context and narrative depth essential for a complete interpretation. Consequently, a synergistic integration of both methods, where visual insights guide textual exploration and vice versa, offers the most robust pathway to a comprehensive assimilation of knowledge.

An online encyclopedia for learning digital architecture concepts could facilitate easier understanding of complex knowledge connections for users and potentially integrate into architectural education. Keeping pace with the constantly evolving dynamic nature of digital architecture concepts, innovative methods utilizing information visualization tools can enhance the understanding and tracking of these concepts compared to slower and less efficient traditional knowledge acquisition methods.

The comprehensive, multi-cycle user evaluation process provided decisive empirical validation for the proposed encyclopedia's design and pedagogical efficacy. The strong average System Usability Scale (SUS) score of 75-80 confirmed the platform's high degree of usability and intuitive navigation. More significantly, the mixed-methods approach validated the core theoretical framework of the thesis. The qualitative feedback and quiz results demonstrated that users could effectively distinguish between and leverage the conceptual, procedural, and visual knowledge layers to build a coherent understanding. The observed cognitive challenges with configurational knowledge did not signify a flaw in the model, but rather affirmed the existence of this higher-order, systemic thinking, reinforcing the necessity of its

distinct classification. This finding aligns with principles of Cognitive Load Theory, suggesting that separating complex information into distinct, manageable layers reduces extraneous cognitive load and facilitates deeper processing. Furthermore, eye-tracking analysis revealed diverse exploratory archetypes, providing concrete evidence that the interface successfully supports the intended non-linear, user-driven learning pathways.

These findings demonstrate that the proposed framework functions not merely as a usable tool, but as an effective environment that facilitates the shift from passive information reception to the active, exploratory construction of knowledge, thereby empirically grounding the central claims of this research.

This thesis does not merely present a tool; it proposes a new pedagogical paradigm for digital architecture. It catalyzes a fundamental shift from the passive consumption of linear information to the active construction of knowledge through exploration. The core mechanism driving this shift is the system's explicit, multi-layered classification, situated within a navigable, visual landscape. This approach operationalizes the principles of Exploratory Learning, where the structure of the environment itself becomes a scaffold for inquiry rather than a conduit for predefined answers. This paradigm challenges the traditional, didactic model of education, which often presents knowledge as a finalized, linear narrative. Instead, it embraces the complexity and interconnectedness of digital architecture discourse, arguing that learning in the 21st century requires tools that reflect this networked reality. The proposed encyclopedia, therefore, is an argument in itself: an argument for an educational model that values inquiry over memorization, and relational understanding over isolated facts.

Consequently, this framework transforms the user from a passive recipient into an active architect of their own understanding. This process of discovery and synthesis aligns directly with the core tenets of constructivist theory, empowering learners to build more robust and personalized cognitive frameworks for navigating the complexities of the digital architecture domain. This is where the concepts of digital architecture and constructivist pedagogy intersect. The visual taxonomy developed in this thesis is not just an organizational scheme; it is a cognitive artefact, a tool for thought. By externalizing the relational structure of the domain, it offloads the cognitive burden of organizing vast amounts of information, thereby freeing up

mental resources for higher-order thinking, critical analysis, and creative connection-making. By empirically grounding the claim that a well-designed information architecture through visualization empowers learners, this work lays a foundation for future research into constructivist, technology-enhanced learning environments in architecture and other design-oriented disciplines.



REFERENCES

- Acar, Y.** (2019). Bilginin haritalanması: bilgi, ilişkilendirme ve temsil. *Dosya 42: İlişkisel Bir Eylem Olarak Haritalama*, 19-25.
- Adams, J. L., & Gunn, K. B.** (2012). Digital humanities: where to start. *College & Research Libraries News*, 536-569.
- Aggarwal, C. C., & Zhai, C.** (2012). *Mining Text Data*. New York: Springer Science & Business Media.
- Alexander, P., & Judy, J.** (1988). The interaction of domain-specific and strategic knowledge in academic performance. *Review of Educational Research*, 58(4), 375-404.
- Amadiou, F., Tricot, A., & Mariné, C.** (2009). Prior knowledge in learning from a non-linear electronic document: disorientation and coherence of the reading sequences. *Computers in Human Behavior*, 381-388.
- Anderson, J. R.** (1976). *Language, Memory and Thought*. Hillsdale: NJ: Lawrence Erlbaum.
- Asheim, B., & Coenen, L.** (2005). Knowledge bases and regional innovation systems: comparing nordic clusters. *Research Policy*, 34, 1173-1190.
- Aydınlı, S.** (2016). Tasarım eğitiminde yapılandırıcı paradigma: ‘öğrenmeyi öğrenme’. *Tasarım + Kuram*, 11(20), 1-18.
- Biggs, J., & Tang, C.** (2011). *Teaching for Quality Learning at University*. Glasgow: McGraw-Hill Education.
- Blei, D. M.** (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77 - 84.
- Blei, D. M., Ng, A. Y., & Jordan, M. I.** (2003). Latent dirichlet allocation. *Journal of Machine Learning Research* 3, 993-1022.
- Blumberg, P.** (2009). Maximizing learning through course alignment and experience with different types of knowledge. *Innovative Higher Education (Springer)*, 34, 93-103.
- Bradley, R., & Swartz, N.** (1979). *Possible Worlds: An Introduction to Logic and Its Philosophy*. USA: Hackett Publishing Company.

- Bruner, J.** (1961). The act of discovery. *Harvard Educational Review*, 31(1), 21–32.
- Bümen, N. T.** (2006). Program geliřtirmede bir dönüm noktası: yenilenmiş bloom taksonomisi. *Education and Science*, 3-14.
- Caetano, I., & Leitão, A.** (2020). Architecture meets computation: an overview of the evolution of computational design approaches in architecture. *Architectural Science Review*, 63(2), 165-174.
- Celani, G., Sperling, D. M., & Franco, J. M.** (2015). CAAD conferences: a brief history. *The Next City: Proceedings of the 16th International Conference CAAD Futures 2015*. 47-58. Biblioteca Central Cesar Lattes.
- Chen, C.** (2006). CiteSpace II: detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology*, 359-377.
- Chiu, M.-L., Lin, C.-J., Jeng, T.-S., & Lee, C.-H.** (2002). Re-searching the research problems with CAAD: data Mining in i-CAADRIA. *Proceedings of the 7th International Conference on Computer Aided Architectural Design Research in Asia*, 31-38. Cyberjaya, Malaysia.
- Collins, A. M., & Quillian, M. R.** (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behaviour*, 240-247.
- Collins, A., & Halverson, R.** (2009). *Rethinking Education in the Age of Technology: The Digital Revolution and the Schools*. New York: Teachers College Press.
- Cook, M. P.** (2006). Visual representations in science education: the influence of prior knowledge and cognitive load theory on instructional design principles. *Science Education*, 90(6), 1073-1091.
- Cook, P.** (2022). *D3 Start to Finish: Learn How to Make a Custom Data Visualisation Using D3.js*. Lean Publishing.
- D'Alembert, & Diderot, D.** (1751). *Encyclopedia, or a Systematic Dictionary of the Sciences, Arts, and Crafts (Encyclopédie, ou Dictionnaire raisonné des sciences, des arts et des métiers)*.
- Das, A., Foged, I. W., Jensen, M. B., & Hansson, M. N.** (2019). Collaborative robotic masonry and early stage fatigue prediction. *Architecture in the Age of the 4th Industrial Revolution: Proceedings of the 37th eCAADe and 23rd SIGraDi Conference*, 3, (pp. 171-178). University of Porto.
- Deleuze, G., & Guattari, F.** (1987). *A Thousand Plateaus: Capitalism and Schizophrenia*. University of Minnesota Press.
- Dodge, M., Kitchin, R., & Perkins, C.** (2011). Introducing the map reader. In M. Dodge, R. Kitchin, & C. Perkins, *The Map Reader: Theories of Mapping*

Practice and Cartographic Representation. (pp. 331–344). John Wiley & Sons.

Duda, R. O., & Hart, P. E. (1973). *Pattern Classification and Scene Analysis*. New York: John Wiley & Sons.

Eppler, M., & Burkhard, R. (2004). Knowledge visualization: towards a new discipline and its fields of application.

Feldman, R., & Dagan, I. (1995). Knowledge discovery in textual databases (KDT). *KDD'95: Proceedings of the First International Conference on Knowledge Discovery and Data Mining*, 112 - 117.

Gaikwad, S. V., Chaugule, A., & Patil, P. (2014). Text mining methods and techniques. *International Journal of Computer Applications*, 85(17), 42-45.

Gero, J. S., & Maher, M. L. (1993). *Modeling Creativity and Knowledge-Based Creative Design*. Hillsdale: Routledge.

Giblett, R. (2009). *Literary Theory and Criticism*. Retrieved February 26, 2024, from <https://literariness.org/2017/04/26/the-philosophical-concept-of-rhizome/>

Goldstein, E. B. (2008). *Cognitive Psychology: Connecting Mind, Research, and Everyday Experience*. Belmont: Thomson Wadsworth Publishing Company.

Grootendorst, M. (2022). BERTopic: neural topic modeling with a class-based TF-IDF procedure. *ArXiv*, abs/2203.05794.

Güleç, G. (2021). Mimarlıkta temsil ve tasarım araçları olarak haritalar. *Grid Mimarlık, Planlama ve Tasarım Dergisi*, 53-73.

Gürer, E. (2014). *Tasarımda sayısal düşünmenin fenomenolojisi*. (Ph.D. Thesis). Istanbul Technical University, Graduate School of Natural and Applied Sciences, Istanbul.

Güzelci, O. Z., Alaçam, S., & Güzelci, H. (2019). Trend topics and changing concepts of computational design in the last 16 years a content analysis. *Education and Research in Computer Aided Architectural Design in Europe*, (pp. 423-430).

Hartigan, J. (1975). *Clustering Algorithms*. New York: John Wiley and Sons.

Hearst, M. (2003). *What is Text Mining?* Retrieved February 23, 2024, from <https://people.ischool.berkeley.edu/~hearst/text-mining.html>

Hillier B., H. J. (1987). Ideas are in things: an application of the space syntax method to discovering house genotypes. *Environment and Planning B: Planning and Design*, 363-385.

- Hillier, B.** (2007). *Space is the Machine: A Configurational Theory of Architecture*. London: Space Syntax.
- Hillier, B.** (2014). Space syntax as a method and as a theory. *21st International Seminar on Urban Form*. Porto.
- Hippner, H., & Rentzmann, R.** (2006). *Informatik Spektrum*, Springer Verlag.
- Hotho, A., Nürnberger, A., & Paaß, G.** (2005). A brief survey of text mining. *Journal for Language Technology and Computational Linguistics*, 20(1), 19–62.
- Hovestadt, L., Hirschberg, U., & Fritz, O.** (2020). *Atlas of Digital Architecture: Terminology, Concepts, Methods, Tools, Examples, Phenomena*. Basel: Birkhäuser.
- Isomenger, I.** (2018). Digital humanities and transdisciplinary practice: towards a rigorous conversation. *Transdisciplinary Journal of Engineering & Science*, 116-138.
- ITU MBL.** (2025). *Architectural Design Computing Program*. Retrieved February 14, 2025, from <https://mbl.itu.edu.tr/program-overview/>
- Jain, A. K.** (2010). Data clustering: 50 years beyond k-means. *Pattern Recognition Letters*, 651–666.
- Jencks, C.** (1971). Charles Jencks's Map of Evolutionary Architecture. Retrieved March 29, 2024, from <https://www.jencksfoundation.org/explore/image/evolutionary-tree-to-the-year-2000>.
- Jo, T.** (2019). *Text Mining: Concepts, Implementation, and Big Data Challenge*. Springer International Publishing AG.
- Jonassen, D., Beissner, K., & Yacci, M.** (1993). *Structural Knowledge: Techniques for Representing, Conveying, and Acquiring Structural Knowledge*. Hillsdale: Lawrence Erlbaum Associates.
- Jong, T. d., & Ferguson-Hessler, M.** (1996). Types and qualities of knowledge. *Educational Psychologist*, 31(2), 105-113.
- Kogut, B., & Zander, U.** (1992). Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization Science*, 3(3), 383-397.
- Kolarevic, B.** (2000). Digital architectures. *Proceedings of the 20th Annual Conference of the Association for Computer Aided Design in Architecture (ACADIA)*, (pp. 251-256).
- Koutamanis, A.** (2000). Digital architectural visualization. *Automation in Construction*, 9(4), 347-360.

- Luo, Y., Pang, P. C.-I., & Chang, S.** (2025). Enhancing exploratory learning through exploratory search with the emergence of large language models. *Proceedings of the 58th Hawaii International Conference on System Sciences, HICSS 2025*. 44-53. Honolulu: IEEE Computer Society.
- Lynch, N.** (2022). Symbolic knowledge structures and intuitive knowledge structures.
- Martens, B., & Turk, Ž.** (2003). Cumulative index of CAAD: current status and future directions. *International Journal of Architectural Computing*, 1(2), 219–231.
- Martens, B., Brown, A., & Turk, Z.** (2003). Automated classification of CAAD-related publications: conditions for setting-up a keywording system. *Proceedings of the Annual Conference of the Association for Computer Aided Design in Architecture 2003*. (pp. 364–369).
- McClelland, J. L., & Rogers, T. T.** (2003). The parallel distributed processing approach to semantic cognition. *Nature*, 310-322.
- McCormick, R.** (1997). Conceptual and procedural knowledge. *International Journal of Technology and Design Education*, 141-159.
- Menken, Q.** (2019). *Constellating images: bilderatlases as a tool to develop criticality towards visual culture*. (Master Education in Arts). Willem de Kooning Academy at Rotterdam University of Applied Sciences (RUAS), Piet Zwart Institute, Rotterdam.
- Millar, A.** (2000). The scope of perceptual knowledge. *Philosophy* 75.
- Millar, A.** (2019). *Knowing by Perceiving*. England: Oxford University Press.
- Miner, G., Elder, J., Hill, T., Nisbet, R., Delen, D., & Fast, A.** (2012). *Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications*. USA: Academic Press (Elsevier).
- Njoo, M. K.** (1994). *Exploratory learning with a computer simulation: learning processes and instructional support*. (Ph.D. Thesis). Eindhoven University of Technology, Industrial Engineering and Innovation Sciences, Eindhoven.
- Norman, D.** (1993). *Things That Make Us Smart: Defending Human Attributes in the Age of the Machine*. Addison-Wesley.
- Oxford Learner's Dictionaries.** (2024). Chair, *Oxford Learner's Dictionaries*. Retrieved February 17, 2024 from https://www.oxfordlearnersdictionaries.com/definition/english/chair_1?q=chair

- Öztemiz, S., & Özel, N.** (2020). Dijital insani bilimler araçları üzerine bir değerlendirme. *Ankara Üniversitesi Dil ve Tarih-Coğrafya Fakültesi Dergisi*, 60(1), 390-414.
- Pathak, S., & Pathak, S.** (2020). Data Visualization Techniques, Model and Taxonomy. In J. Hemanth, M. Bhatia, & O. Geman, *Data Visualization and Knowledge Engineering: Spotting Data Points with Artificial Intelligence* (pp. 249-271). Switzerland: Springer.
- Piaget, J.** (2008). Intellectual evolution from adolescence to adulthood. *Human Development*, 51(1), 40–47.
- Po, L., Bikakis, N., Desimoni, F., & Papastefanatos, G.** (2020). Principles of Data Visualization. In Y. Ding, P. Groth, & J. Hendler, *Linked Data Visualization: Techniques, Tools, and Big Data* (pp. 19-45). Morgan & Claypool Publishers.
- Polanyi, K.** (1963). Ports of trade in early societies. *The Journal of Economic History*, 23(1), 30-45.
- Price, D. J.** (1965). Networks of scientific papers. *Science*, 510-515.
- RealEye.** (2024). RealEye eye tracking, Retrieved December 12, 2024, from www.realeye.io.
- RealEye.** (2025). RealEye eye tracking, Retrieved January 02, 2025, from www.realeye.io.
- Reif, F.** (1987). Instructional design, cognition, and technology: applications to the teaching of scientific concepts. *Journal of Research in Science Teaching*, 24(4), 309-324.
- Reif, F., & Allen, S.** (1992). Cognition for interpreting scientific concepts: a study of acceleration. *Cognition and Instruction*, 9(1), 1-44.
- Russell, B.** (2020). *Stanford Encyclopedia of Philosophy: A Priori Justification and Knowledge*. Retrieved April 18, 2024, from <https://plato.stanford.edu/entries/apriori/>
- S.Zhao, Wang, J., Ji, J., & Eko, A.** (2024). Proximity or alienation? Can knowledge type influence the relationship between proximity and enterprise innovation performance? *Technological Forecasting & Social Change* 202.
- Salton, G., & Buckley, C.** (1988). Term-weighting approaches in automatic text. *Information Processing & Management*, 24, 513–523.
- Schoenfeld, A. H.** (1985). *Mathematical Problem Solving*. Orlando: Academic Press.

- Sedig, K., Parsons, P., Dittmer, M., & Haworth, R.** (2014). Human-Centered Interactivity of Visualization Tools: Micro- and Macro-level Considerations. *Handbook of Human Centric Visualization*, (pp. 717-743). Springer.
- Selvi, K.** (2011). Phenomenological Approach in Education. In Tymieniecka, AT. (Eds.), *Analecta Husserliana* (Vol. 95). Dordrecht: Springer.
- Shafiabady, N., Lee, L., Rajkumar, R., Kallimani, V., Akram, N. A., & Isa, D.** (2016). Using unsupervised clustering approach to train the support vector machine for text classification. *Neurocomputing*, 4-10.
- Soliman, S., Taha, D., & Sayad, Z. E.** (2019). Architectural education in the digital age computer applications: between academia and practice. *Alexandria Engineering Journal, Elsevier*, 809-818.
- Solomon, K. O., Medin, D. L., & Lynch, E.** (1999). Concepts do more than categorize. *Trends in Cognitive Science*, 99-105.
- Sweller, J., Ayres, P., & Kalyuga, S.** (2011). *Cognitive Load Theory*. New York: Springer.
- Tabah, A. N.** (1999). Literature dynamics: studies on growth, diffusion, and epidemics. *Annual Review of Information Science and Technology*, 34, 249-286.
- Talib, R., Hanif, M. K., Ayesha, S., & Fatima, F.** (2016). Text mining: techniques, applications and issues. *International Journal of Advanced Computer Science and Applications*,, 414-418.
- Thomas, J. J., & Cook, K. A.** (2005). *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. National Visualization and Analytics Center.
- Tonta, Y.** (2012). Bilgi Sınıflama, Bilgi Düzenleme ve Bilgi Erişim. In Ö. Külcü, T. Çakmak, & N. Özel, *Prof. Dr. K. Gülbün Baydur'a armağan* (pp. 155-172).
- Turk, Z., Cerovsek, T., & Martens, B.** (2001). The topics of CAAD: a machine's perspective. *Proceedings of the Ninth International Conference on Computer Aided Architectural Design Futures*. 547-560. Eindhoven.
- Url-1** <<https://visuwords.com/knowledge>>, date retrieved 09.01.2023.
- Url-2** <<https://www.visualthesaurus.com>>, date retrieved 09.01.2023.
- Url-3** <<https://www.lexipedia.com>>, data retrieved 09.01.2023.
- Url-4** <<https://kumu.io/altspaces/learning-architecture-da#map-Ub9U5s8J>>, data retrieved 19.01.2025.
- Url-5** <<http://cluster.cis.drexel.edu/~cchen/citespace>>, data retrieved 09.01.2023.

Url-6 <<https://graphcommons.com>>, data retrieved 09.01.2023.

Url-7 <<https://www.vosviewer.com>>, data retrieved 09.01.2023.

Url-8 <<https://observablehq.com/@d3/gallery>>, data retrieved 20.02.2023.

Url-9 <<https://papers.cumincad.org/about.html>>, data retrieved 02.10.2023.

Vercruysse, S., & Kuiper, M. (2020). Intuitive representation of computable knowledge.

Ware, C. (2013). *Information Visualization: Perception for Design*. USA: Morgan Kaufmann, Elsevier.

Wikipedia. (2022). *Wikipedia*. Retrieved October 12, 2022, from <https://en.wikipedia.org/wiki/Wikipedia>

Witten, I. H. (2004). Text Mining. In M. P. Singh, *The Practical Handbook of Internet Computing*. USA: Chapman and Hall/CRC.

APPENDICES

APPENDIX A1: Heatmap results from group 1 on the zoomable sunburst (RealEye, 2024).

APPENDIX A2: Heatmap results from group 1 on the conceptual knowledge (RealEye, 2024).

APPENDIX A3: Heatmap results from group 1 on the procedural knowledge (RealEye, 2024).

APPENDIX A4: Heatmap results from group 1 on the visual knowledge (RealEye, 2024).

APPENDIX A5: Heatmap results from group 1 on the configurational knowledge (RealEye, 2024).

APPENDIX B1: Heatmap results from group 2 on the zoomable sunburst (RealEye, 2025).

APPENDIX B2: Heatmap results from group 2 on the conceptual knowledge (RealEye, 2025).

APPENDIX B3: Heatmap results from group 2 on the procedural knowledge (RealEye, 2025).

APPENDIX B4: Heatmap results from group 2 on the visual knowledge (RealEye, 2025).

APPENDIX B5: Heatmap results from group 2 on the configurational knowledge (RealEye, 2025).

APPENDIX C1: Heatmap results from group 3 on the zoomable sunburst (RealEye, 2025).

APPENDIX C2: Heatmap results from group 3 on the conceptual knowledge (RealEye, 2025).

APPENDIX C3: Heatmap results from group 3 on the procedural knowledge (RealEye, 2025).

APPENDIX C4: Heatmap results from group 3 on the visual knowledge (RealEye, 2025).

APPENDIX C5: Heatmap results from group 3 on the configurational knowledge (RealEye, 2025).

APPENDIX D: Survey, interview questions, AI quiz, SUS Questionnaire (approved by the ethics committee).

APPENDIX E: Pseudo code for RIS to JSON conversion and preprocessing.

APPENDIX F: Pseudo code for hierarchical clustering of JSON data.

APPENDIX A1

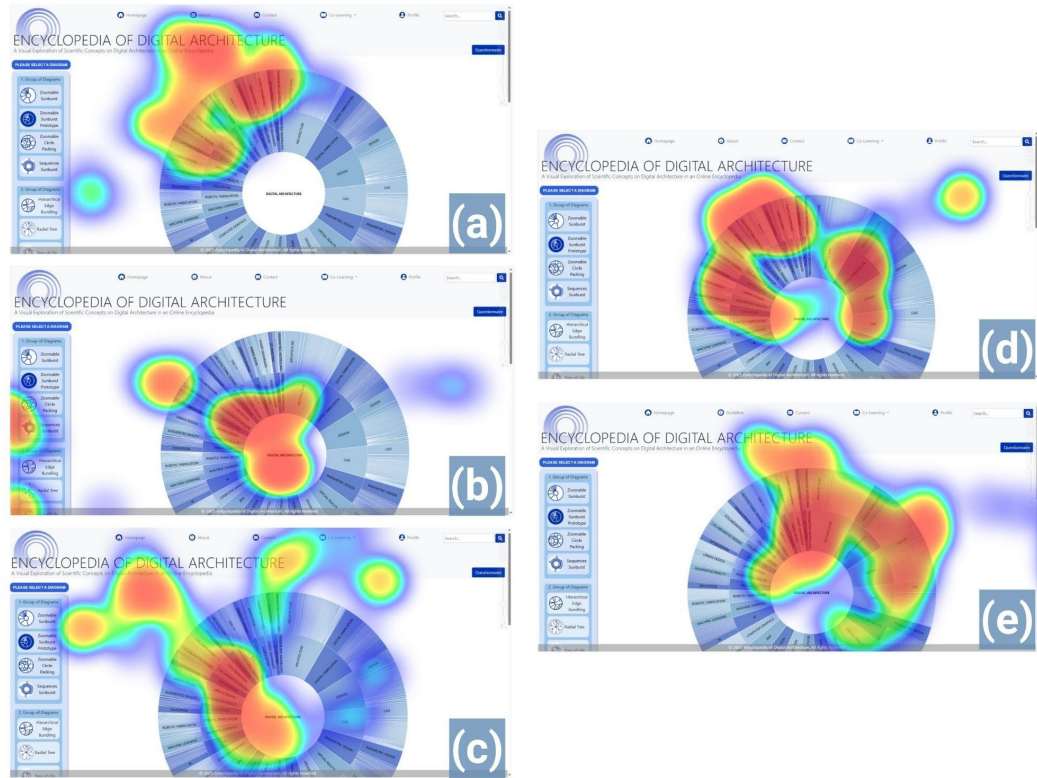


Figure A.1.1 : Heatmap results from group 1 on the zoomable sunburst: (a)First participant. (b)Second participant. (c)Third participant. (d)Fourth participant. (e)Fifth participant (RealEye, 2024).



APPENDIX A2

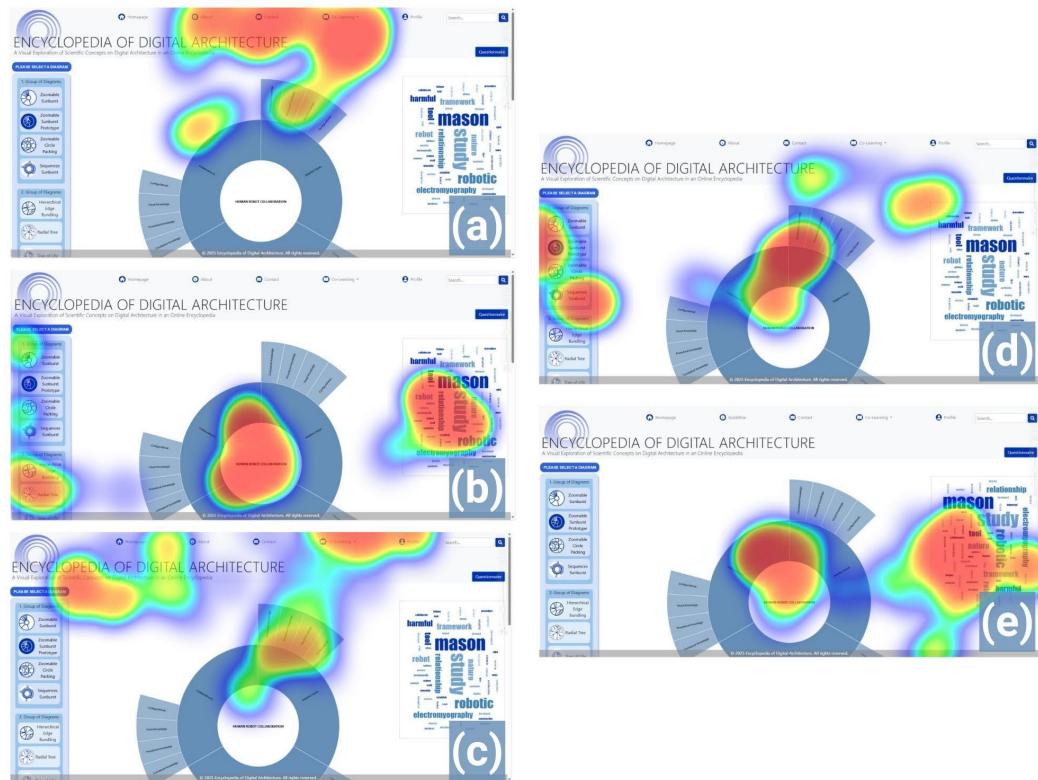


Figure A.2.1 : Heatmap results from group 1 on the conceptual knowledge: (a)First participant. (b)Second participant. (c)Third participant. (d)Fourth participant. (e)Fifth participant (RealEye, 2024).



APPENDIX A3

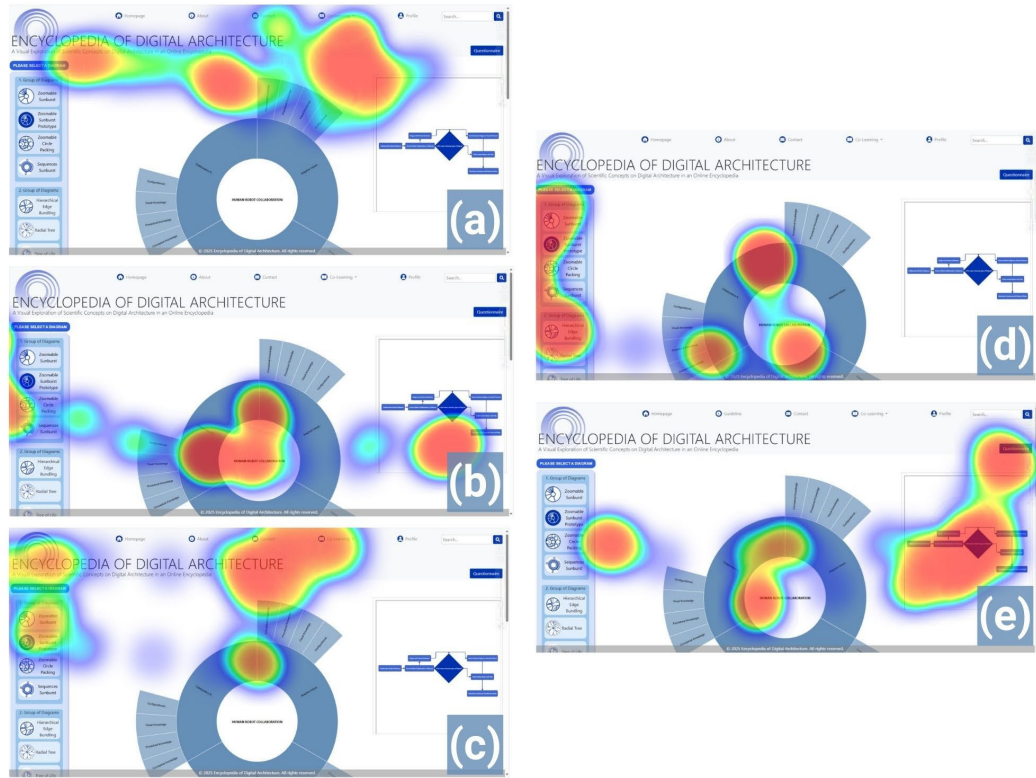


Figure A.3.1 : Heatmap results from group 1 on the procedural knowledge: (a)First participant. (b)Second participant. (c)Third participant. (d)Fourth participant. (e)Fifth participant (RealEye, 2024).



APPENDIX A4

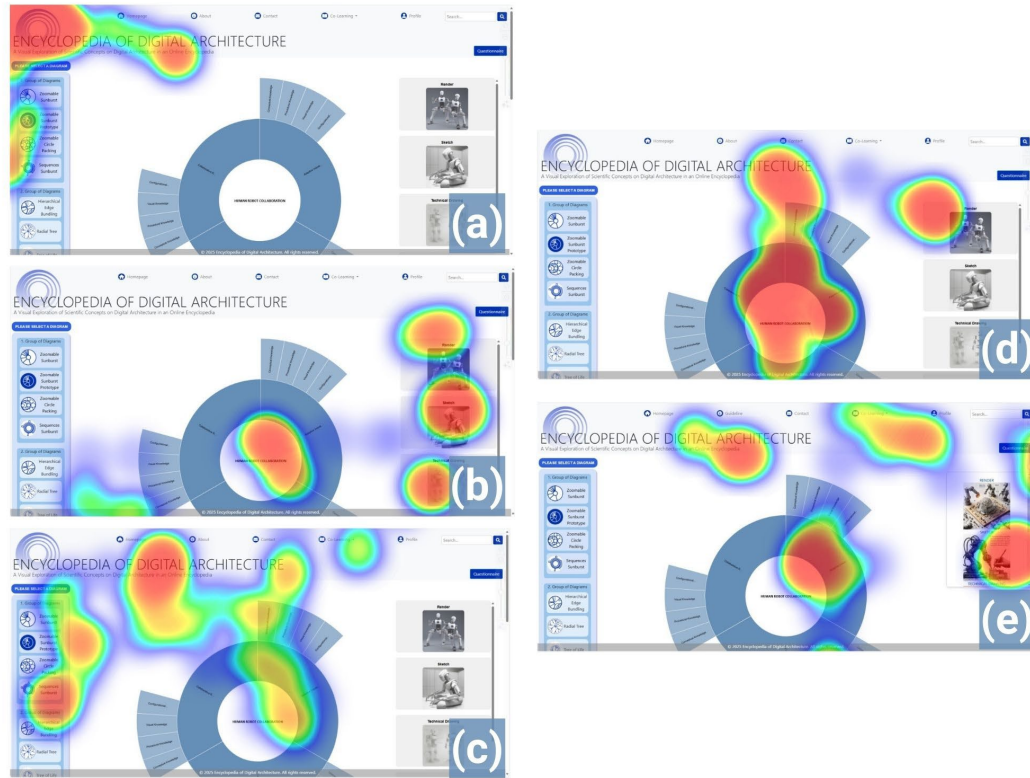


Figure A.4.1 : Heatmap results from group 1 on the visual knowledge: (a)First participant. (b)Second participant. (c)Third participant. (d)Fourth participant. (e)Fifth participant (RealEye, 2024).



APPENDIX A5

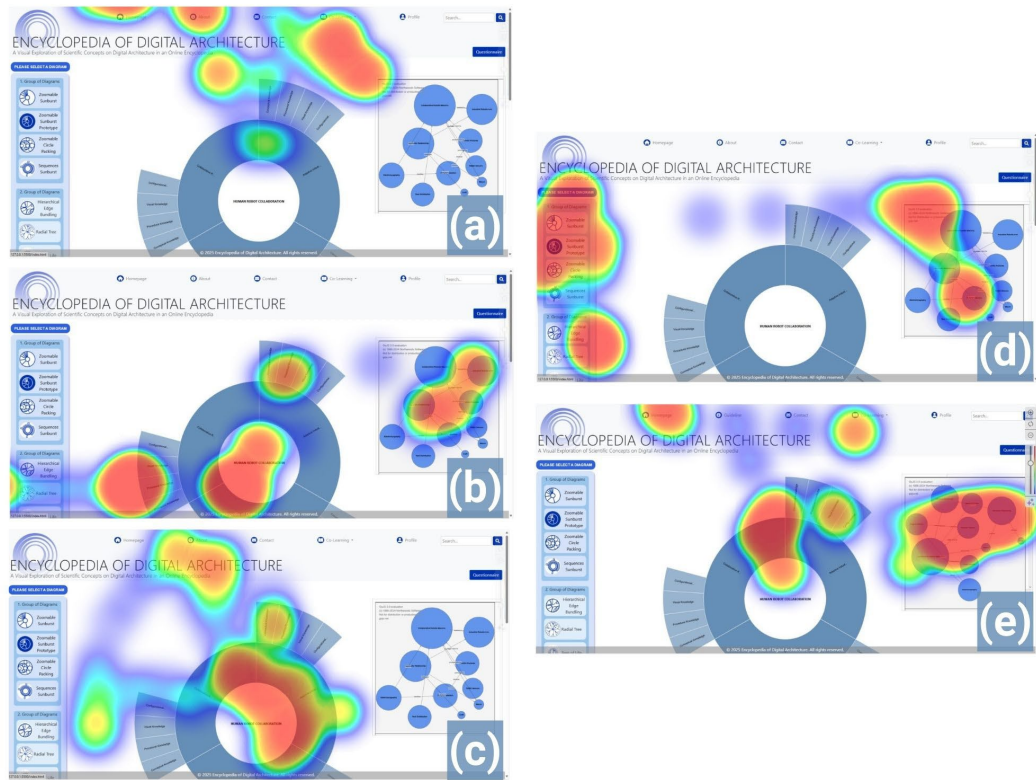


Figure A.5.1 : Heatmap results from group 1 on the configurational knowledge:
 (a)First participant. (b)Second participant. (c)Third participant. (d)Fourth participant.
 (e)Fifth participant (RealEye, 2024).



APPENDIX B1

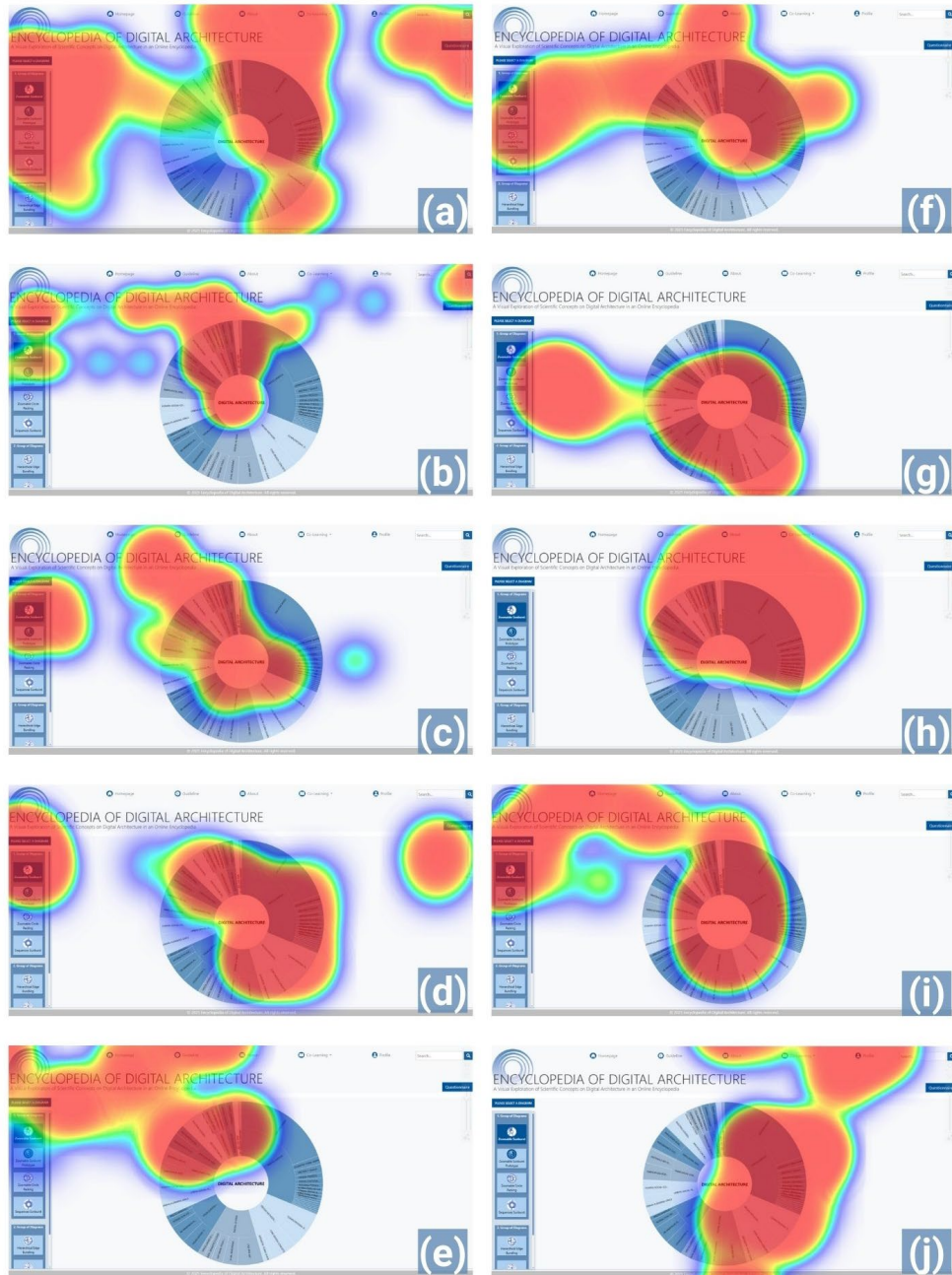


Figure B.1.1 : Heatmap results from group 2 on the zoomable sunburst: (a)First participant. (b)Second participant. (c)Third participant. (d)Fourth participant. (e)Fifth participant. (f)Sixth participant. (g)Seventh participant. (h)Eighth participant. (i)Ninth participant. (j)Tenth participant (RealEye, 2024).



APPENDIX B2

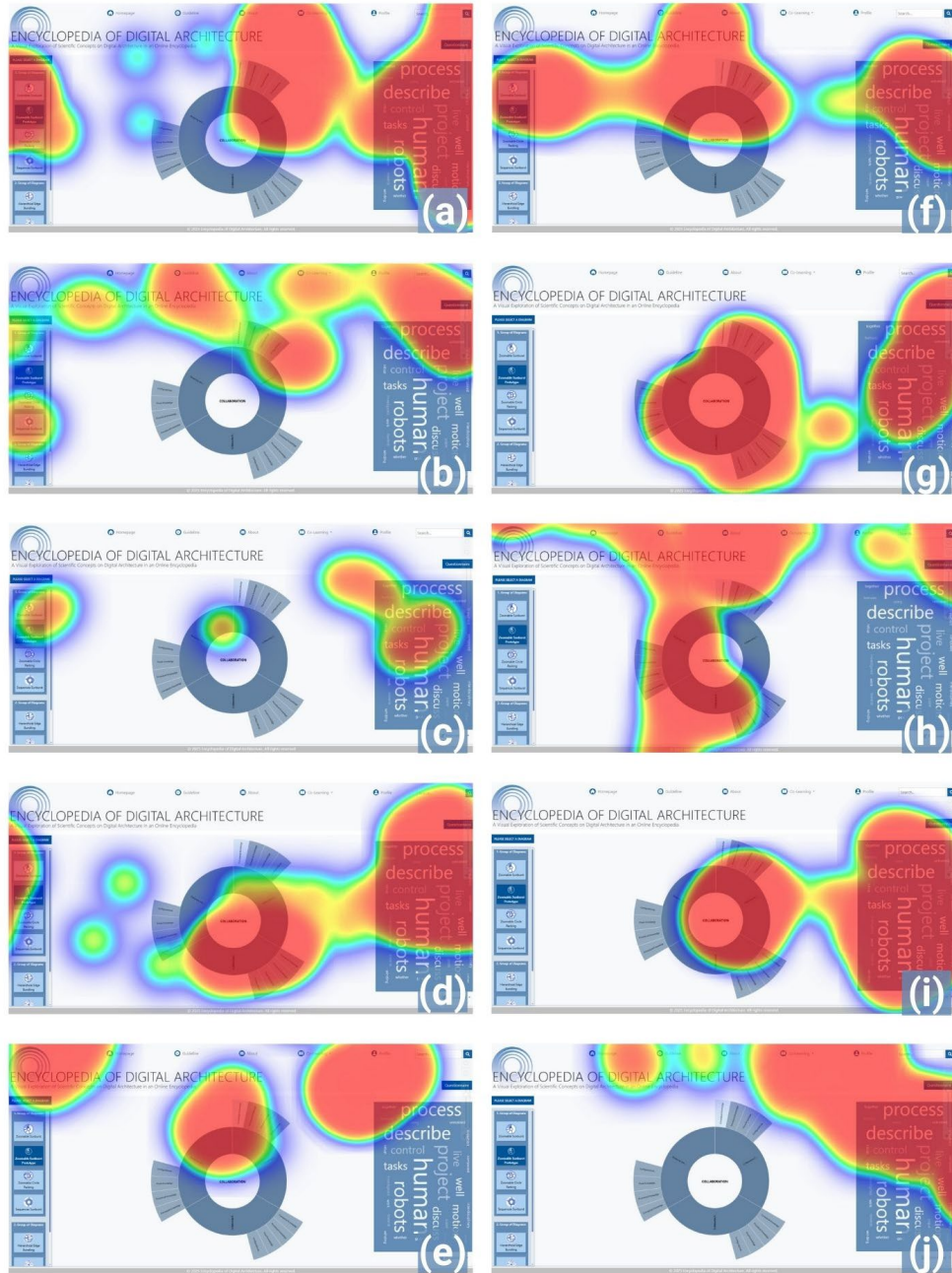


Figure B.2.1 : Heatmap results from group 2 on the conceptual knowledge: (a)First participant. (b)Second participant. (c)Third participant. (d)Fourth participant. (e)Fifth participant. (f)Sixth participant. (g)Seventh participant. (h)Eighth participant. (i)Ninth participant. (j)Tenth participant (RealEye, 2024).



APPENDIX B3

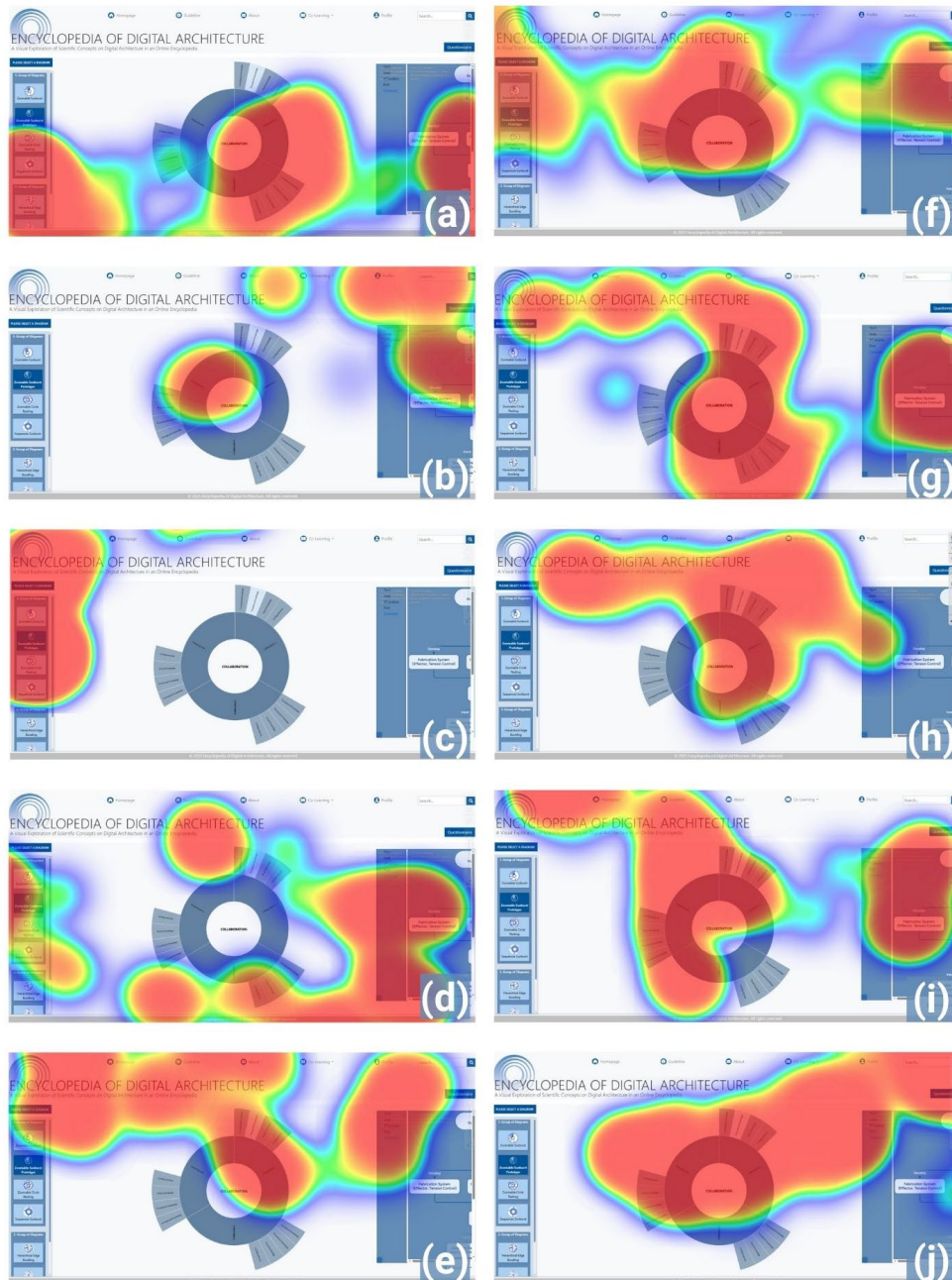


Figure B.3.1 : Heatmap results from group 2 on the procedural knowledge: (a)First participant. (b)Second participant. (c)Third participant. (d)Fourth participant. (e)Fifth participant. (f)Sixth participant. (g)Seventh participant. (h)Eighth participant. (i)Ninth participant. (j)Tenth participant (RealEye, 2024).



APPENDIX B4

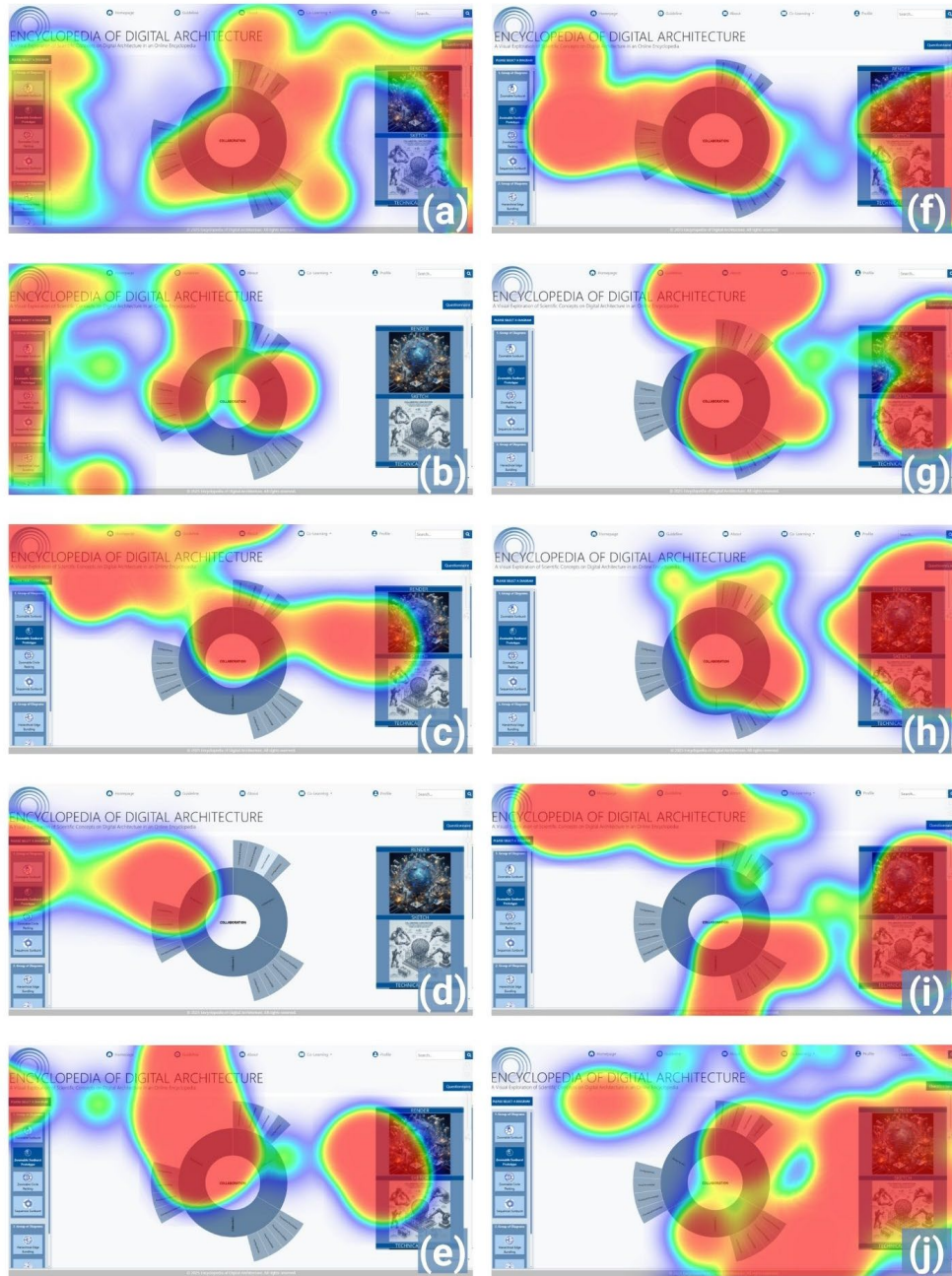


Figure B.4.1 : Heatmap results from group 2 on the visual knowledge: (a)First participant. (b)Second participant. (c)Third participant. (d)Fourth participant. (e)Fifth participant. (f)Sixth participant. (g)Seventh participant. (h)Eighth participant. (i)Ninth participant. (j)Tenth participant (RealEye, 2024).



APPENDIX B5

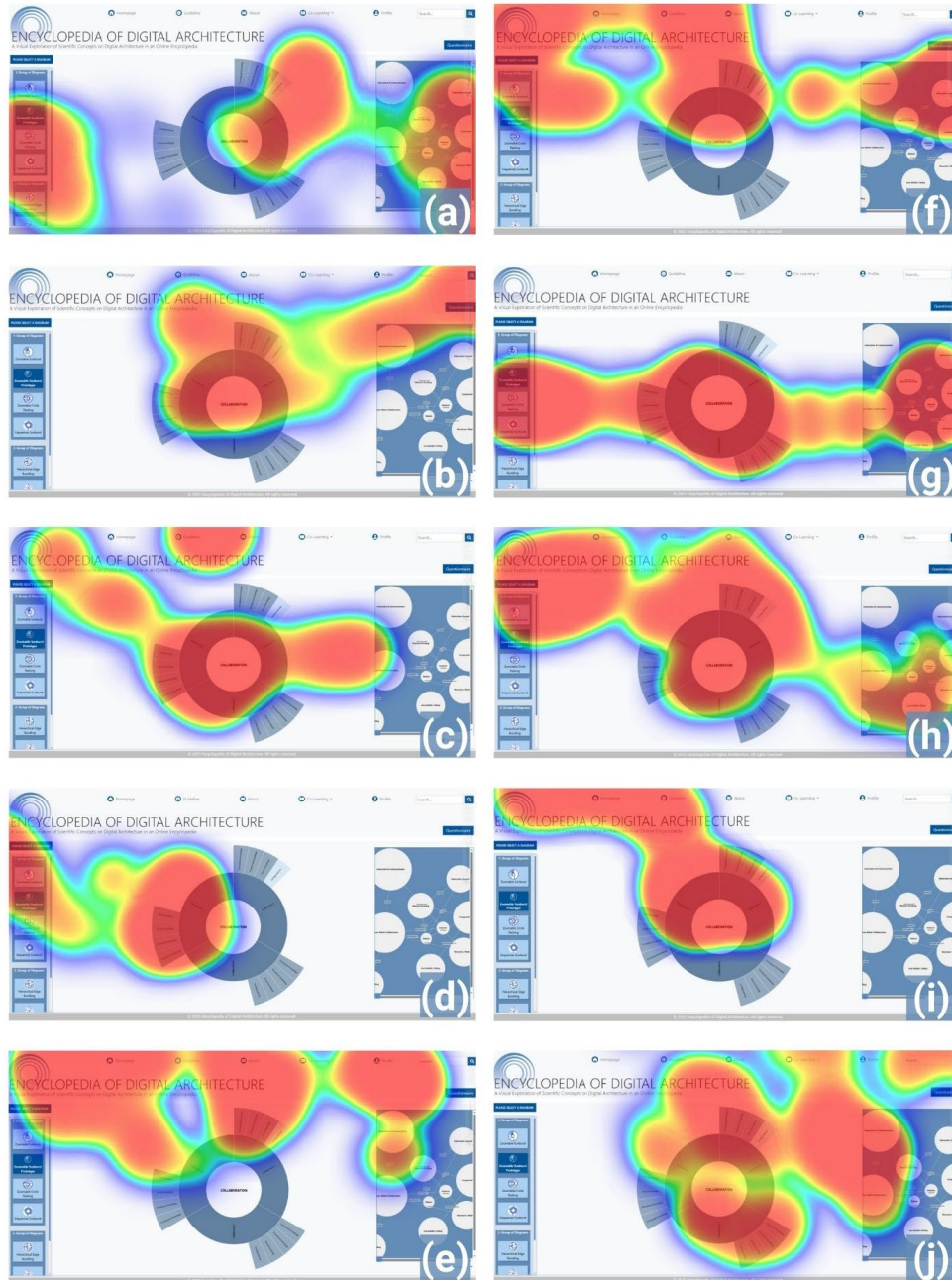


Figure B.5.1 : Heatmap results from group 2 on the configurational knowledge:
 (a)First participant. (b)Second participant. (c)Third participant. (d)Fourth participant.
 (e)Fifth participant. (f)Sixth participant. (g)Seventh participant. (h)Eighth participant.
 (i)Ninth participant. (j)Tenth participant (RealEye, 2024).



APPENDIX C1

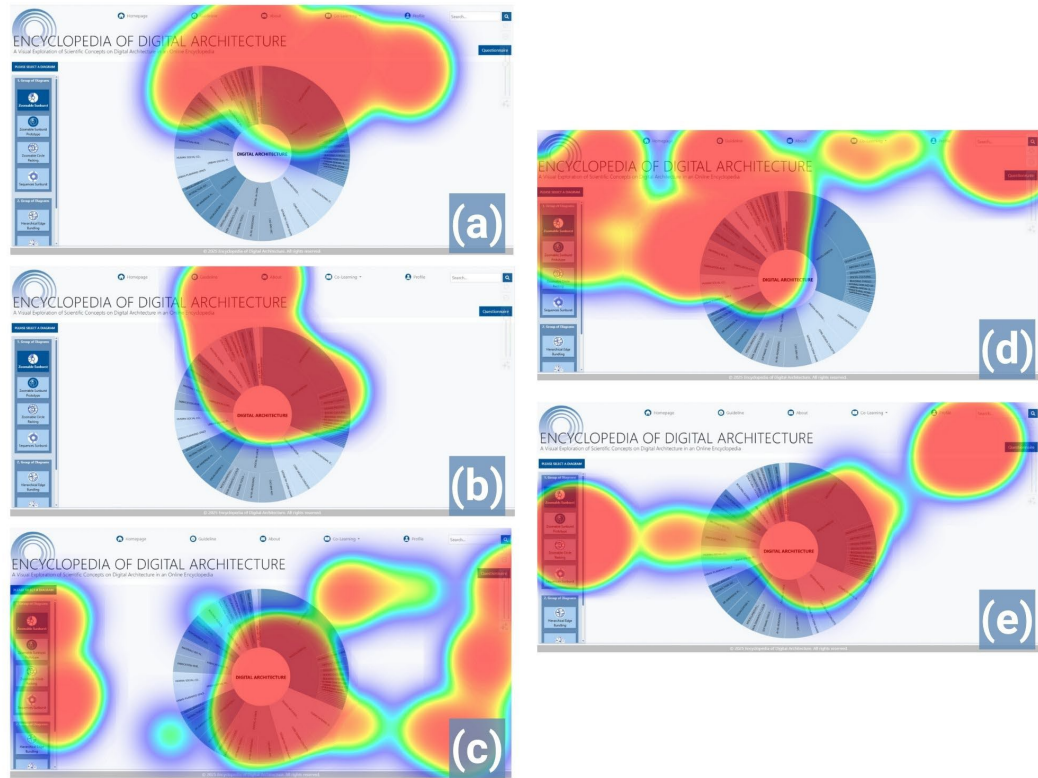


Figure C.1.1 : Heatmap results from group 3 on the zoomable sunburst: (a)First participant. (b)Second participant. (c)Third participant. (d)Fourth participant. (e)Fifth participant (RealEye, 2024).



APPENDIX C2



Figure C.2.1 : Heatmap results from group 3 on the conceptual knowledge: (a)First participant. (b)Second participant. (c)Third participant. (d)Fourth participant. (e)Fifth participant (RealEye, 2024).



APPENDIX C3

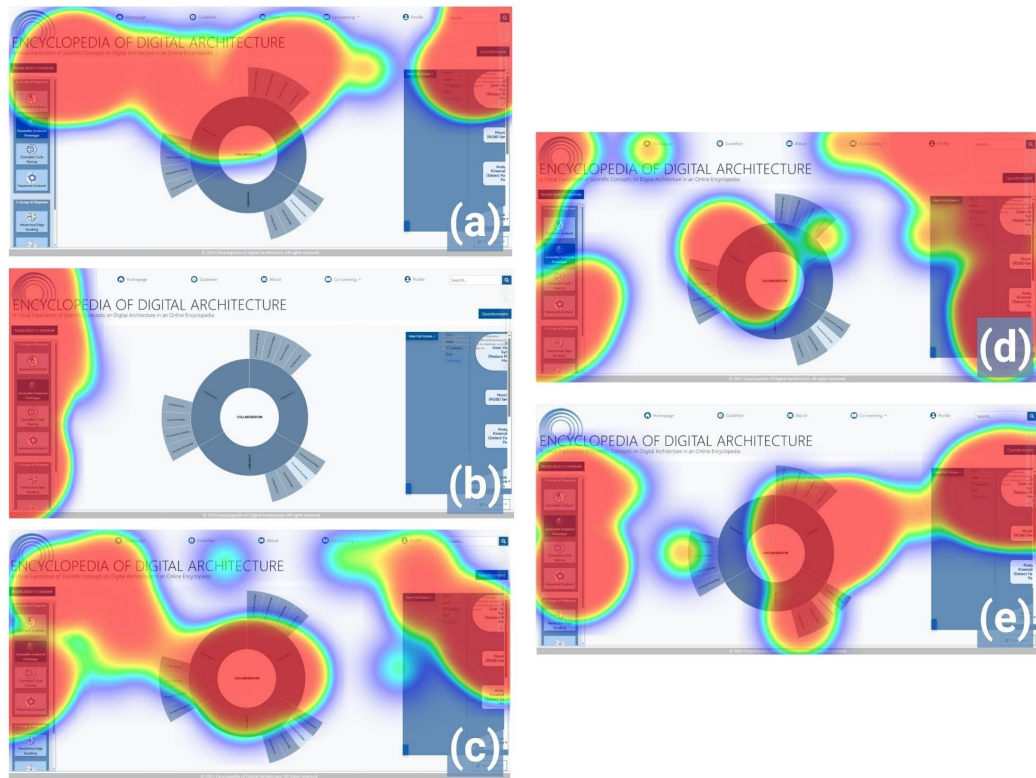


Figure C.3.1 : Heatmap results from group 3 on the procedural knowledge: (a)First participant. (b)Second participant. (c)Third participant. (d)Fourth participant. (e)Fifth participant (RealEye, 2024).



APPENDIX C4



Figure C.4.1 : Heatmap results from group 3 on the visual knowledge: (a)First participant. (b)Second participant. (c)Third participant. (d)Fourth participant. (e)Fifth participant (RealEye, 2024).



APPENDIX C5

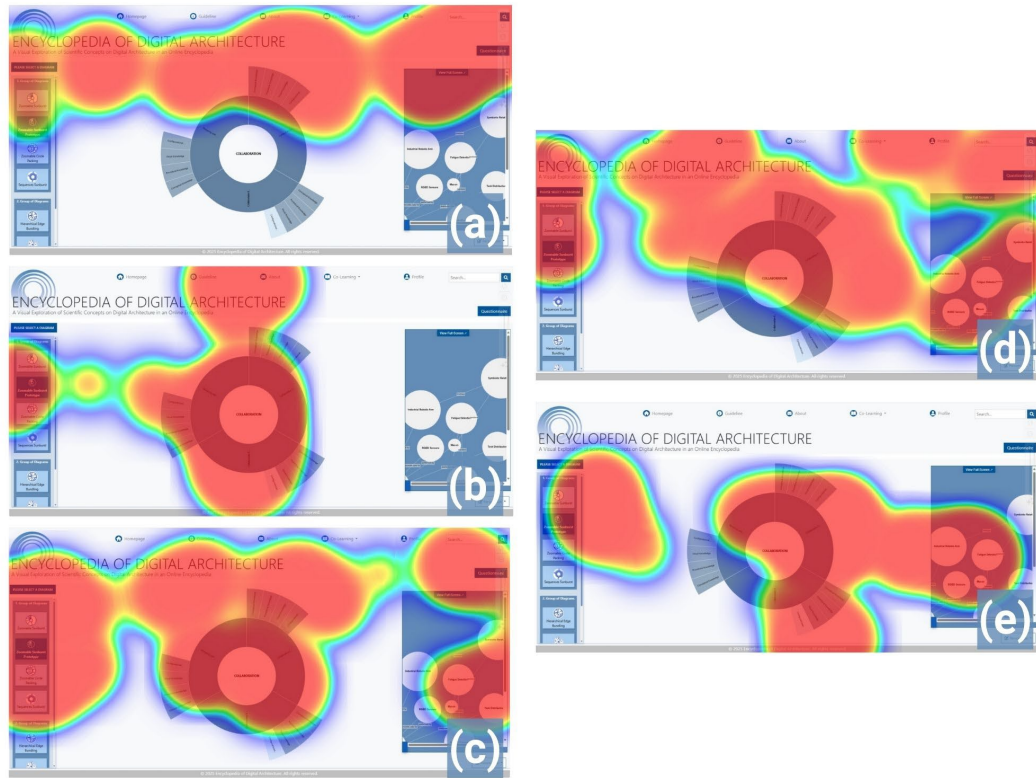


Figure C.5.1 : Heatmap results from group 3 on the configurational knowledge:
(a)First participant. (b)Second participant. (c)Third participant. (d)Fourth participant.
(e)Fifth participant (RealEye, 2024).



APPENDIX D

1. Kullanım Öncesi Bilgilendirme Metni

- *Çalışma kapsamında ses kaydı ve ekran kaydı alınacak; göz takip yazılımı aracılığı ile görev tamamlama süresi, göz hareket örüntüleri, odak noktaları, ısı haritaları, fare hareketleri, duyuşsal veriler (mutluluk, şaşkınlık, nötr yüz ifadesi) gibi verilerden bir veya birkaçı toplanacaktır. Çalışmanın süresi yaklaşık olarak 30 dakikadır.*

1. Sistem Hakkında Genel Bilgi:

- Sistem Tanımı: Bu sistem, Dijital Mimarlık alanında bilgiye daha hızlı ve etkili bir şekilde erişmenizi sağlamak için tasarlanmıştır. Hiyerarşik bir yapı içinde, belirli bir konuya dair kavramsal, prosedürel, görsel ve konfigürasyonel bilgiyi size sunar.
- Amaç: Sistem, akademik makalelerdeki terimlere aşinalık kazandırmayı, kavramsal ilişkileri ve metodolojik süreçleri daha iyi anlamanızı, görsel olarak hatırlamanızı ve bu bilgiyi yeni durumlara uygulamanızı sağlamayı hedefler.

2. Sistem Bileşenleri ve Bilgi Katmanları:

- Kavramsal Bilgi Katmanı: Kavramlar ve bu kavramlar arasındaki ilişkiler hakkında bilgi verir. Bu katman, konuyu anlamınıza yardımcı olacak temel kavramları tanıtır.
- Prosedürel Bilgi Katmanı: Belirli bir süreci ya da metodolojiyi aşama aşama öğrenmenizi sağlar. Akış şemaları bu katmanda yer alır.
- Görsel Bilgi Katmanı: Konuyla ilgili görsel içerikler sunar.
- Konfigürasyonel Bilgi Katmanı: Sistemin size sunduğu bilgilere ek olarak, kendi konfigürasyonel bilginizi ekleyerek sistemi kişiselleştirmenizi sağlar. Bu katmanda, yeni ilişkiler oluşturabilir ve mevcut bilgiyi kendi bakış açınıza göre düzenleyebilirsiniz.

3. Sistem Kullanımına Yönelik Talimatlar:

- Hiyerarşik Yapı: Sistem hiyerarşik bir yapıya sahiptir. Yani, ana başlıklardan alt başlıklara doğru makalelere ulaşabileceksiniz. Her bir başlık, daha detaylı bilgi katmanlarına ulaşmanıza olanak tanır.
- Navigasyon: Sistemde gezinirken, konuyla ilgili alt başlıklara tıklayarak daha detaylı bilgilere erişebilirsiniz. İlgilendiğiniz konuyu seçtikten sonra, ilgili bilgi katmanları karşınıza çıkacaktır.

- Geri Bildirim: Sistemi kullanırken herhangi bir zorlukla karşılaşırsanız veya herhangi bir konuda geri bildirimde bulunmak isterseniz, lütfen bunları ifade ediniz.

4. Görevlerin Amacı:

- Size vereceğimiz görevler, sistemin sağladığı hiyerarşik bilgi yapılarını kullanarak belirli bir konuyu ne kadar iyi kavradığınızı ve bu bilgiyi ne derece etkili bir şekilde uygulayabildiğinizi ölçmeyi amaçlamaktadır.

5. Beklentiler:

- Görevleri yerine getirirken, adımları dikkatli bir şekilde takip etmeniz ve öğrendiğiniz bilgileri hiyerarşik yapıyı kullanarak daha derinlemesine anlamaya çalışmanız beklenmektedir.
- Görev sonrasında, öğrendiklerinizi ne kadar iyi kavradığınızı ve hangi alanlarda zorlandığınızı belirten geri bildirimlerde bulunmanızı ve yapay zekâ ile üretilmiş bir mini quizi tamamlamanızı isteyeceğiz.

2. Göz Takip Verisi

(Deneyisel, Davranışsal, Gözlemsel Veri, Kullanıcı Deneyimi Araştırması):

Aşağıdaki size verilen görevleri gerçekleştirirken sizden bir göz izleme yazılımıyla aşağıdaki verilerden bir veya birkaçı toplanacaktır:

Görev Tamamlama Süresi

Göz Hareket Örüntüleri

Odak Noktaları

Isı Haritaları

Fare Hareketleri

Duyusal Veriler (mutluluk, şaşkınlık, nötr yüz ifadesi)

3. Görev Basamakları

SİSTEMİN KULLANICIYA SAĞLADIĞI FAYDALARI ÖLÇME

1. Lütfen soldaki diyagram sekmesinde bulunan diyagramlardan en üstteki Zoomable Sunburst diyagramını seçiniz.

2. Diyagramda bulunan “Robotic Fabrication” konusunun alt katmanlarını inceleyiniz. Daha sonra “Human-Robot Collaboration” katmanına tıklayınız.

3. Human-Robot Collaboration katmanının alt katmanları arasında bulunan “Collaborative Robotic Masonry and Early Stage Fatigue Prediction” makalesini seçiniz.
4. “Collaborative Robotic Masonry and Early Stage Fatigue Prediction” makalesinin alt katmanlarından Conceptual Knowledge katmanını seçiniz.
5. Conceptual Knowledge katmanının alt katmanlarını inceleyiniz ve aklınızda kalan en az 2 ana katmanı ve bunların alt katmanlarından 3 katmanı not alınız.
6. Dairenin ortasına tıklayıp geri gelerek Procedural Knowledge katmanında bulunan akış şemasını inceleyiniz ve yorumlayınız. Akış şemasını en fazla 5 aşama ile özetleyiniz. Sizce bu çalışmanın amacı nedir? Bu katmandan Visual Knowledge katmanına geçmeden önce aklınızda bu çalışmaya dair bir görsel canlandı mı? Görselden bahsedebilir misiniz, mümkünse çizer misiniz?
7. Visual Knowledge katmanında bulunan görselleri inceleyiniz. Hangi görselin sizin için daha anlamlı olduğunu ve bunun nedenini belirtiniz. Bu görseller makalenin bütününe dair bilgi edinmenize yardımcı oldu mu? Sizce bir önceki katmanlardan aldığınız bilgiyle örtüşüyor mu?
8. Configurational Knowledge katmanında bulunan ağ haritalamayı inceleyiniz, konuyla ilgili çağrışımsal olarak aklınıza gelen 3 fikri anahtar kelimelerle ilgili düğümlere ekleyiniz.

MINI AI QUIZ

Kullanıcının Kavramsal Bilgi Anlayışını Ölçme

Çalışmanın ana odağı nedir?

- A) Duvarcılık inşaatının tarihini araştırmak
- B) Ofis ortamlarındaki zararlı durumları incelemek
- C) Bir insan ile mekanik olarak eklemli bir araç arasında simbiyotik bir ilişki geliştirmek
- D) İnşaat projeleri üzerinde hava koşullarının etkisini analiz etmek

Kullanıcının Prosedürel Bilgi Anlayışını Ölçme

Zararlı durum ve kas gerilimini incelemek için hangi prosedür kullanılmıştır?

- A) X-ray görüntüleme
- B) Termografi
- C) Yüzey elektromiyografi
- D) Ultrason taraması

Kullanıcının Görsel Bilgi Anlayışını Ölçme

Çalışmada kullanılan teknoloji hangisidir?

- A) Sanal gerçeklik başlıkları
- B) Uydu iletişim sistemleri
- C) RGBD sensörler ve yüzey elektromiyografi
- D) 3D baskı teknolojisi

Kullanıcının Konfigürasyonel Bilgi Anlayışını Ölçme

Çalışmada görevlerin duvar işçisi ve robot arasında dağıtılmasının amacı nedir?

- A) Yorgunluğu ve zararlı duruşları tespit etmek
- B) İşçiler arasındaki rekabeti artırmak
- C) İnsan iş gücü ihtiyacını azaltmak
- D) Robot kolunun verimliliğini test etmek

4. Sistem Kullanılabilirlik Ölçeği-System Usability Score (SUS)

(Kullanıcıların belirli bir sistem veya arayüzle etkileşimini genel olarak değerlendirmek için)

(Beşli likert ölçeği>Kesinlikle katılmıyorum (1)—Kesinlikle katılıyorum(5))

- 1-Bu sistemi daha sık kullanmak istiyorum.
- 2-Bu sistemin olması gerekenden daha karmaşık olduğunu düşünüyorum.
- 3-Sistemin basit ve kullanımı kolay olduğunu düşünüyorum.
- 4-Bu sistemi kullanmak için teknik desteğe ihtiyacım var.
- 5-Sistemin sorunsuz çalıştığını ve iyi entegre edildiğini görüyorum.
- 6-Bence sistemde çok fazla düzensizlik var.
- 7-Bence çoğu insan bu sistemi hızlı bir şekilde öğrenebilir.
- 8-Bu sistemin zaman alıcı olduğunu düşünüyorum.
- 9-Bu sistemi kullanırken kendimi güvende hissediyorum.
- 10-Bu sistemi kullanmaya başlamadan önce öğrenmem gereken çok şey olduğunu düşünüyorum.

APPENDIX E

```
FUNCTION ConvertRISToJSON(ris_file_content)
  articles_array = SPLIT ris_file_content by "ER -" delimiter
  json_output_array = []

  FOR EACH article_text in articles_array
    raw_article = PARSE article_text into key-value pairs (e.g., "T1" -> "Title")

    // Preprocessing and Filtering
    IF raw_article["N2"] has less than 100 words THEN
      CONTINUE
    END IF

    // Create Clean JSON Object
    processed_article = {}
    processed_article["Author(s)"] = raw_article["AU"]
    processed_article["Year"] = CLEAN(raw_article["PY"])
    processed_article["Title"] = raw_article["T1"]
    processed_article["Keywords"] =
    NORMALIZE_KEYWORDS(raw_article["KW"])
    processed_article["Summary"] = raw_article["N2"]

    ADD processed_article to json_output_array
  END FOR

  RETURN json_output_array
END FUNCTION
```



APPENDIX F

```
FUNCTION CreateHierarchicalCluster(json_data)
// 1. Initialization and Keyword Extraction
root_node = CREATE_NODE("Digital Architecture")
all_keywords = new Map()

FOR EACH article in json_data
// Normalize and split keywords (e.g., "3D MODELING, CAD" -> ["3D
MODELING", "CAD"])
keywords = NORMALIZE_AND_SPLIT(article["Keywords"])

// 2. Build Keyword Tree
FOR EACH keyword_phrase in keywords
// Create a path of nodes from the root for multi-word phrases
// e.g., "ROBOTIC FABRICATION" becomes root -> ROBOTIC ->
FABRICATION
current_node = root_node
words = SPLIT keyword_phrase by space
FOR EACH word in words
current_node = GET_OR_CREATE_CHILD_NODE(current_node, word)
END FOR

// Attach article title to the final node of the path
ADD_ARTICLE_TITLE(current_node, article["Title"])
END FOR
END FOR

// 3. Refinement and Categorization (Post-processing)
// Group related terms under a single canonical form
CONSOLIDATE_SYNONYMS(root_node) // (e.g., "MODELING" and
"MODELLING" -> "MODELING")

// Assign nodes to predefined main categories and sub-categories
CATEGORIZE_NODES(root_node)

// Group single-article nodes for clarity
AGGREGATE_SINGLE_NODES(root_node)

// 4. Final Output
RETURN root_node // The root of the final hierarchical JSON structure
END FUNCTION
```



CURRICULUM VITAE

Name Surname : Esranur KARACİF

EDUCATION:

- **B.Sc.** : 2015, Vienna University of Technology, Architecture
- **M.A.** : 2019, Anadolu University, Visual Communication Design
- **M.Sc.** : 2020, Istanbul Technical University, Architectural Design
- **M. Eng.** : 2022, Hacettepe University, Construction Technologies
- **M.Sc.** : 2025, Ankara University, Informatics

PUBLICATIONS AND PRESENTATIONS ON THE THESIS:

- **Karacif, E., Gürer, E.** 2025. Dijital mimarlık bilgisinin tasnif ve temsili: Bir çevrim içi ansiklopedi modeli önerisi, Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi, 40(3), 1509-1524. <https://doi.org/10.17341/gazimmfd.1488572>
- **Karacif, E., Gürer, E.** 2024. Hierarchical Information Extraction and Structured Navigation in Architectural Design Computing: Leveraging Machine Learning for Enhanced Research Discovery, 3rd International Graduate Research Symposium (IGRS'24), 53.

OTHER PUBLICATIONS AND PRESENTATIONS:

- **Karacif, E., Yazici, S.** 2023. A Methodology for Material-Based Computational Design Supported by Mobile Augmented Reality Application, Architecture and Planning Journal (APJ), 28(3), <https://doi.org/10.54729/2789-8547.1221>
- **Karacif, E., Alkışer Bregger, Y.** 2023. Sosyal Konut Yarışmalarının Mekânsal Kalite Bağlamında Değerlendirilmesi: Viyana Modeli, Konut ve Yaşam Çevreleri, Ed. Ö. Burcu Özdemir Sarı, İdealkent Press.
- **Karacif, E., Gürer, E.** 2022. A Decision Support System Proposal on the Usage of Extended Reality SDKs in AEC Disciplines, Eskişehir Technical University Journal of Science and Technology A - Applied Sciences and Engineering, 23, 17-30. <https://doi.org/10.18038/estubtda.1165368>.
- **Karacif, E.** 2022. Tasarım, Uygulama, Araç İlişkisinde Endüstri 4.0 Devriminden Endüstri 2.0 Devrimine Mimarlıkta Hesaplamalı Tasarım, MSTAS 2022 (Mimarlıkta Sayısal Tasarım 16. Ulusal Sempozyumu).
- **Karacif, E., Alkışer Bregger, Y.** 2022. Mekânsal Kalite Bağlamında Viyana'da Sosyal Konut Yarışmaları, 7. Kent Araştırmaları Kongresi (ICUS 2022), 429.