

Semantische Dokumentenindexierung mit generativer KI

Dr. Dimitri Busch
Fraunhofer-Informationszentrum Raum und Bau
(IRB)
Stuttgart, Deutschland

Prof. Dr. Dmytro Lande
Nationale Technische Universität
„Kiewer Polytechnisches Institut Igor Sikorsky,,
Kiew, Ukraine

Einführung

- In diesem Vortrag handelt es sich um die Anwendung der großen Sprachmodelle zur semantischen Indexierung von Dokumenten im Bereich Architekturgeschichte.
- Die Dokumente (Datensätze) enthalten Referenzinformationen über Zeitschriftenartikel, wie z.B. Titel, Kurzreferate (Abstracts), Deskriptoren (Schlüsselwörter).
- Unter den großen Sprachmodellen (Large Language Models, LLM) werden Systeme verstanden, die in der Lage sind, Text basierend auf Eingabe und Kontext zu generieren.
- Die Indexierung ist die Zuordnung von Deskriptoren zu Dokumenten zum Zweck der inhaltlichen Beschreibung und des Wiederauffindens (Retrieval) dieser Dokumente.

Semantische Indexierung

Bei der semantischen Indexierung werden zu Dokumenten nicht nur Deskriptoren sondern auch Relationen zwischen den Deskriptoren zugeordnet.

Beispieldokument:

Titel: 1, 2, 3, 6: Early Gothic Architecture and Perfect Numbers.

Abstract: Knowledge of numbers should not be despised [...] We are instructed in number to avoid confusion. Take away number in all things, and they all perish. Take away computation from the world, and all things are encompassed by blind ignorance; people who are ignorant of the knowledge of reckoning cannot be distinguished from the other animals'. Thus wrote Isidore of Sevilla in the 7th century (Book III, 4). Number clearly mattered. Indeed, the medieval world seems to have taken to heart the words expressed in the Book of Wisdom (11.21): 'Omnia in mensura et numero et pondere fecisti'. In such a number-obsessed world, one would expect number to play a significant role in the design of medieval architecture, for without number, a building, like everything else, would perish. And indeed, many medieval descriptions of buildings display a great concern with numerical values. After a short introduction into the issues involved, this paper will concentrate on the use of perfect numbers as a design principle in several select examples

[den Hartog, E., 2014, Architectural Histories, 2(1), <http://doi.org/10.5334/ah.bu>]

Relationen: Deskriptor-Paaren

- Gothic Architecture; Perfect Numbers;
- Medieval World; Book of Wisdom;
- Medieval Descriptions; Numerical Values;

...

Verfahren zur semantischen Indexierung mithilfe eines LLM

Erstelle einen Dokumentensatz

Für jedes Dokument

Erzeuge ein Prompt für ein LLM, um Paare verwandter Deskriptoren zu generieren

Führe Sie das Prompt aus

Schreibe die Deskriptor-Paare in eine CSV-Datei

Konvertiere die CSV-Datei in eine SVG-Datei (Graph, Semantisches Netzwerk)

Beispieldokument

Titel: 1, 2, 3, 6: Early Gothic Architecture and Perfect Numbers

Abstract: ‘Knowledge of numbers should not be despised [...] We are instructed in number to avoid confusion. Take away number in all things, and they all perish. Take away computation from the world, and all things are encompassed by blind ignorance; people who are ignorant of the knowledge of reckoning cannot be distinguished from the other animals’. Thus wrote Isidore of Sevilla in the 7th century (Book III, 4). Number clearly mattered. Indeed, the medieval world seems to have taken to heart the words expressed in the Book of Wisdom (11.21): ‘Omnia in mensura et numero et pondere fecisti’. In such a number-obsessed world, one would expect number to play a significant role in the design of medieval architecture, for without number, a building, like everything else, would perish. And indeed, many medieval descriptions of buildings display a great concern with numerical values. After a short introduction into the issues involved, this paper will concentrate on the use of perfect numbers as a design principle in several select examples

[den Hartog, E., 2024, Architectural Histories 2(1), <http://doi.org/10.5334/ah.bu>]

Prompt für ein LLM

Eine Eingabeaufforderung (Prompt) besteht aus einer Anfrage gefolgt von einem Text. Dieser Text enthält den Titel, das Abstract und gegebenenfalls die Deskriptoren, die zum Dokument bereits zugeordnet wurden.

Beispielprompt:

Output 20 pairs of the most related entities from the text. Each entity should be described in no more than 3 words. Format of the output: 'entity 1; entity 2'. Each pair is displayed on a new line. Text:

1, 2, 3, 6: Early Gothic Architecture and Perfect Numbers. Knowledge of numbers should not be despised

Ausgabe für ein Prompt

Prompt:

Output 20 pairs of the most related entities from the text. Each entity should be described in no more than 3 words. Format of the output: 'entity 1; entity 2'. Each pair is displayed on a new line. Text:

1, 2, 3, 6: Early Gothic Architecture and Perfect Numbers. Knowledge of numbers should not be despised

Ausgabe:

Gothic Architecture; Perfect Numbers

Numbers; Computation

World; Blind Ignorance

Knowledge; Reckoning

Isidore of Sevilla; 7th century

Medieval World; Book of Wisdom

Design; Architecture

Number; Building

Medieval Descriptions; Numerical Values

Introduction; Design Principle

Examples; Select Examples

... ..

Konvertierung von CSV in SVG mit CSVGraph

(<https://bigsearch.space/uli.html>)

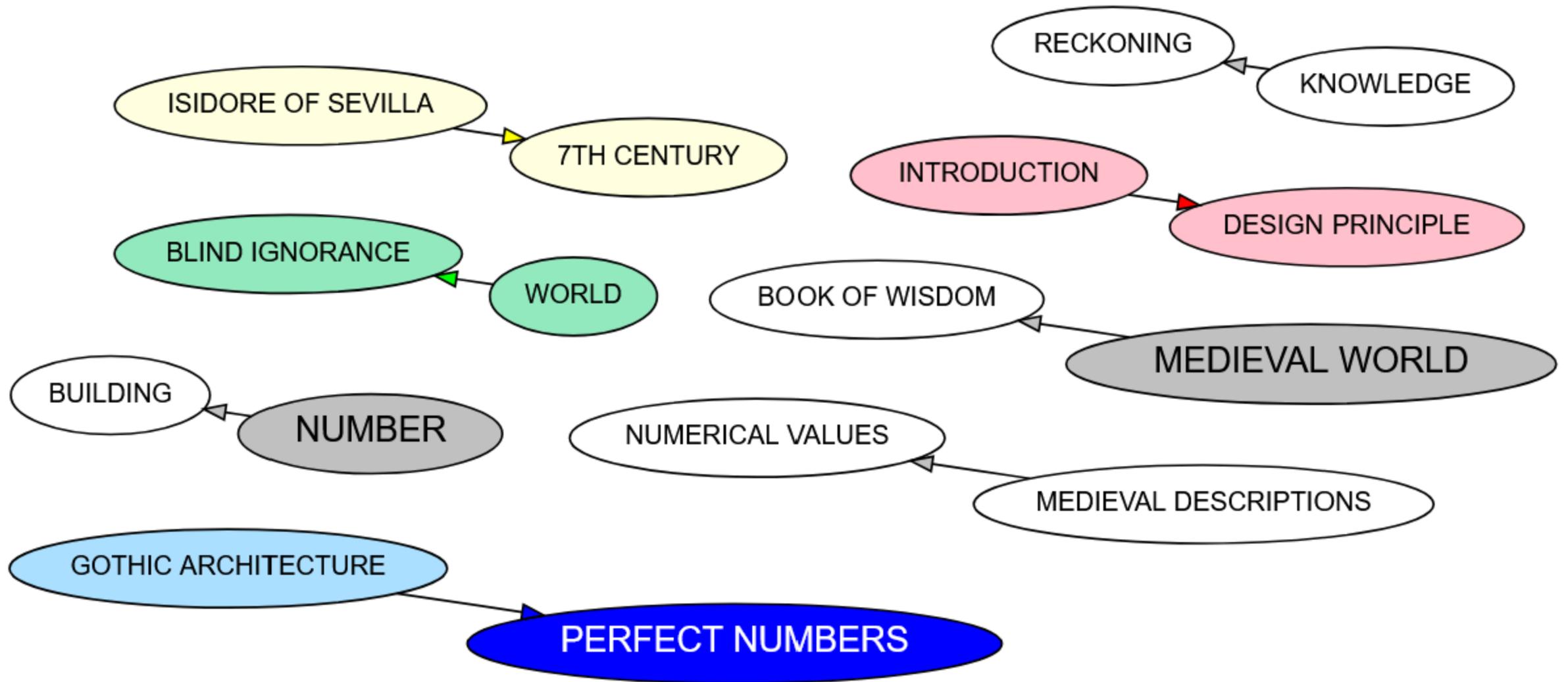


```
Gothic Architecture; Perfect Numbers
Numbers; Computation
World; Blind Ignorance
Knowledge; Reckoning
Isidore of Sevilla; 7th century
Medieval World; Book of Wisdom
Design; Architecture
Number; Building
Medieval Descriptions; Numerical Values
Introduction; Design Principle
Examples; Select Examples
```

Google Google News Bing Bing News Wiki.EN Wiki.UA

Directed Full

Fragment des Semantisches Netzwerkes



Semantische Indexierung: Implementierung

- Erzeugung und Ausführung von Prompts:

Azure OpenAI (FhGenie);

GPT 3.5 Turbo;

Java;

Langchain4j

- Konvertierung von CSV in SVG:

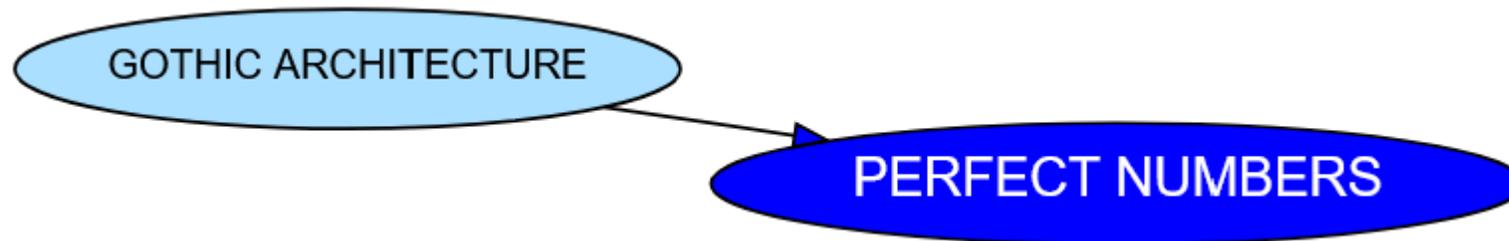
CSVGraph (<https://bigsearch.space/uli.html>)

Prototypisches Suchprogramm

- Zur Implementierung des vorliegenden Modells wurde ein Suchprogramm entwickelt, das nach dem Klicken auf einen Knoten oder eine Kante des semantischen Netzwerks die entsprechenden Dokumente aus der Prototypdatenbank findet und anzeigt.
- Die Suche erfolgt mithilfe einer Abfrage, die diesem Programm über den Linkparameter an einem Knoten oder einer Kante des semantischen Netzwerks übermittelt wird.
- Das Suchprogramm wurde als JSP (Java Server Page, Servlet) implementiert, auf dem Apache Tomcat ausgeführt wird.
- Die Prototypdatenbank wurde als Lucene-Index implementiert
- 39 Dokumente (nur Referenzinformationen) aus der Open-Access-Zeitschrift Architectural Histories, <https://journal.eahn.org/>

Klicken auf den Knoten „Gothic Architecture“

- Wenn man auf einen Knoten klickt, werden Dokumente gesucht, die den entsprechende Deskriptor enthalten.
- **Knoten:** „Gothic Architecture“
- **Link:** <http://betaindex.de/eahweb?q=%22GOTHIC ARCHITECTURE%22>
- **Lucene-Anfrage:** „GOTHIC ARCHITECTURE“



Klicken auf den Knoten „Gothic Architecture“: Treffer

Start of search: 26.11.2024 03:00:52

Query: "GOTHIC ARCHITECTURE"

Hits: 4

ID - 005

TI - 1, 2, 3, 6: Early Gothic Architecture and Perfect Numbers

AB - 'Knowledge of numbers should not be despised [...] We are instructed in number to avoid confusion. Take away number in all things, and they all perish. Take away computation from the world, and all things are encompassed by blind ignorance; people who are ignorant of the knowledge of reckoning cannot be distinguished from the other animals?. Thus wrote Isidore of Sevilla in the 7th century (Book III, 4). Number clearly mattered. Indeed, the medieval world seems to have taken to heart the words expressed in the Book of Wisdom (11.21): 'Omnia in mensura et numero et pondere fecisti?. In such a number-obsessed world, one would expect number to play a significant role in the design of medieval architecture, for without number, a building, like everything else, would perish. And indeed, many medieval descriptions of buildings display a great concern with numerical values. After a short introduction into the issues involved, this paper will concentrate on the use of perfect numbers as a design principle in several select examples.

RN - GOTHIC ARCHITECTURE; PERFECT NUMBERS

RN - NUMBERS; COMPUTATION

RN - WORLD; BLIND IGNORANCE

.....

Klicken auf den Knoten „Gothic Architecture“: Treffer (Fortsetzung)

ID - 021

TI - Reviews Winter 2020

AB - Bergdoll, B. A Review of Miriam Paeslack, *Constructing Imperial Berlin: Photography and the Metropolis*. Minneapolis: University of Minnesota Press, 2019. Fozi, S. A Review of Robert Bork, *Late Gothic Architecture: Its Evolution, Extinction, and Reception*. Turnhout, Belgium: Brepols, 2018. Higgott, A. A Review of Eva Branscome, *Hans Hollein and Postmodernism: Art and Architecture in Austria 1958?1985*. London and New York: Routledge, 2018. Vassallo, J. A Review of Martino Stierli, *Montage and the Metropolis: Architecture, Modernity, and the Representation of Space*. New Haven; London: Yale University Press, 2018; and Craig Buckley, *Graphic Assembly: Montage, Media, and Experimental Architecture in the 1960s*. Minneapolis: University of Minnesota Press, 2019. Wheeler, K. A Review of Conor Lucey, *Building Reputations. Architecture and the Artisan, 1750?1830*. Manchester: Manchester University Press, 2018.

RN - BERGDOLL; MIRIAM PAESLACK

RN - BORK; LATE **GOthic ARCHITECTURE**

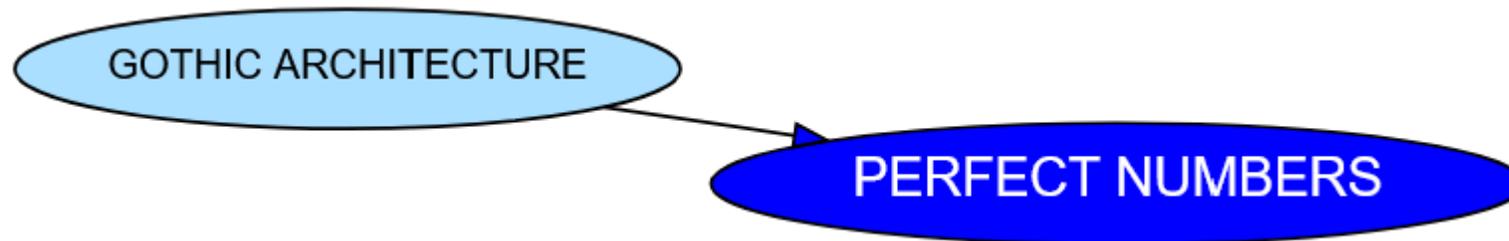
RN - HIGGOTT; EVA BRANSCOME

....

.

Klicken auf eine Kante

- Wenn man auf einer Kante klickt, werden Dokumente gesucht, die das entsprechende Deskriptor-Paar enthalten.
- **Kante:** zwischen „Gothic Architecture“ und „Perfect Numbers“
- **Link:** <http://betaindex.de/eahweb?q=%22GOTHIC ARCHITECTURE%22+%22PERFECT NUMBERS%22>
- **Lucene Anfrage:** „GOTHIC ARCHITECTURE“ AND „PERFECT NUMBERS“



Klicken auf eine Kante: Treffer

Start of search: 26.11.2024 03:25:49

Query: "GOTHIC ARCHITECTURE" AND "PERFECT NUMBERS"

Hits: 1

ID - 005

TI - 1, 2, 3, 6: Early Gothic Architecture and Perfect Numbers

AB - 'Knowledge of numbers should not be despised [...] We are instructed in number to avoid confusion. Take away number in all things, and they all perish. Take away computation from the world, and all things are encompassed by blind ignorance; people who are ignorant of the knowledge of reckoning cannot be distinguished from the other animals?. Thus wrote Isidore of Sevilla in the 7th century (Book III, 4). Number clearly mattered. Indeed, the medieval world seems to have taken to heart the words expressed in the Book of Wisdom (11.21): 'Omnia in mensura et numero et pondere fecisti?. In such a number-obsessed world, one would expect number to play a significant role in the design of medieval architecture, for without number, a building, like everything else, would perish. And indeed, many medieval descriptions of buildings display a great concern with numerical values. After a short introduction into the issues involved, this paper will concentrate on the use of perfect numbers as a design principle in several select examples.

RN - GOTHIC ARCHITECTURE; PERFECT NUMBERS

RN - NUMBERS; COMPUTATION

RN - WORLD; BLIND IGNORANCE

Deutsches Dokument

Titel: Bauforschung an figurierten Gewölben der Spätgotik: Das Beispiel der Georgskirche in Augsburg.

Abstract: Eine Vielzahl gotischer Kirchen in Süddeutschland wurde erst in der ausgehenden Spätgotik eingewölbt, sodass die Kirchen zu diesem Zeitpunkt bereits eine längere Baugeschichte hatten. Dieses Phänomen wirft die Frage auf, wie solche Gewölbe geplant und hergestellt wurden, denn die Baumeister mussten mit Unregelmäßigkeiten der bestehenden Bauwerke rechnen und mit den vorhandenen Konstruktionen zurechtkommen. Die Methoden, die zur Planung und Herstellung von spätgotischen Gewölben genutzt wurden, werden am Beispiel der Georgskirche in Augsburg untersucht. Diese Stiftskirche, die bereits im 12. Jahrhundert bestand, wurde in den 1490er Jahren tiefgreifend umgebaut. Das Langhaus der Basilika wurde vergrößert und dabei vollständig neu errichtet; der Chor hingegen wurde lediglich renoviert und eingewölbt. Damit verkörpern diese beiden Bauteile der Kirche zwei unterschiedliche Bauaufgaben: Das Langhaus steht für das Wölben eines Neubaus, während der Chor das Wölben eines älteren Bestandsbaus repräsentiert.

[Voigts, C., 2015, Architectura, Jg.45, Nr.1, S.45-69]

Indexierung des deutschen Dokuments

Prompt:

Geben Sie 20 Paare der am stärksten verwandten Entitäten aus dem aus. Jede Entität sollte mit nicht mehr als 3 Wörtern beschrieben werden. Format der Ausgabe: „Entität 1; Entität 2“. Jedes Paar wird in einer neuen Zeile angezeigt.

Text:

Bauforschung an figurierten Gewölben der Spätgotik: Das Beispiel der Georgskirche in Augsburg....

Ausgabe:

Georgskirche; Spätgotik

Georgskirche; Augsburg

Georgskirche; Gewölbe

gotische Kirchen; Süddeutschland

gotische Kirchen; Spätgotik

Kirchen; Baugeschichte

Gewölbe; Planung

Gewölbe; Herstellung

Baumeister; Unregelmäßigkeiten

Baumeister; bestehende Bauwerke

.....

Zusammenfassung

Der Einsatz generativer Technologien der künstlichen Intelligenz kann den Arbeitsaufwand und die Kosten der semantischen Indexierung verringern.

Die semantische Indexierung und das semantische Netzwerk können ihrerseits zu einer effektiveren Nutzung und Verbreitung wissenschaftlicher Informationen beitragen, indem sie eine semantische Suche, eine einfache Navigation und Benutzerfreundlichkeit ermöglichen.

Literatur

Dmytro Lande, Leonard Strashnoy. GPT Semantic Networking: A Dream of the Semantic Web – The Time is Now. – Kyiv: Engineering, 2023

Vielen Dank für Ihre
Aufmerksamkeit