Automatische Extraktion von Fachterminologie aus kunsthistorischen Volltexten

Juliane Bredack

SKILL 2014
Studierendenkonferenz Informatik
25. September 2014

Überblick

- Einführung
  - Was sind Mehrwortgruppen?
  - Lingo und die Komponenten
  - Das Reallexikon zur Deutschen Kunstgeschichte
- Praktisches Vorgehen
- Ergebnisse
  - Extraktion von Mehrwortgruppen
- Fazit

Bedeutung von Mehrwortgruppen

- Erweiterung der Suchmöglichkeiten im Information Retrieval
- Bedeutungsdifferenzierung auf der Ebene eines einzelnen Wortes
  - Bank → Deutsche Bank
- Aussagekraft und Deutlichkeit
- Zusammenhängende lexikalische Einheit → besitzen als Ganzes eine Bedeutung

Lingo – das Indexierungssystem

- System zur linguistisch und statistisch basierten automatischen Indexierung

- Wörterbuchgestützte Grundform-identifizierung bzw. –reduktion, Kompositazerlegung, algorithmische und lexikalische Mehrwortgruppenerkenung, Synonymrelationierung

http://lex-lingo.blogspot.de

Programmmodul sequencer

- Algorithmische Identifizierung und Extraktion noch unbekannter Mehrwortgruppen

Automatische Extraktion fachterminologischer Mehrwortgruppen mit Lingo.

<Automatische=[(automatisch/a)]>
<Extraktion=[(extraktion/s)]>
<fachterminologischer=[(fachterminologisch)/a]>
<Mehrwortgruppen=[(mehrwortgruppe/s)]>
<mit=[(mit/c)]>
<Lingo=[(lingo/s)]>
:./PUNC

Programmmodul sequencer

Bildung von Wortmustern, basierend auf Wortklassen → Identifizierung von Wortgruppen

Automatische Extraktion fachterminologischer Mehrwortgruppen mit Lingo.

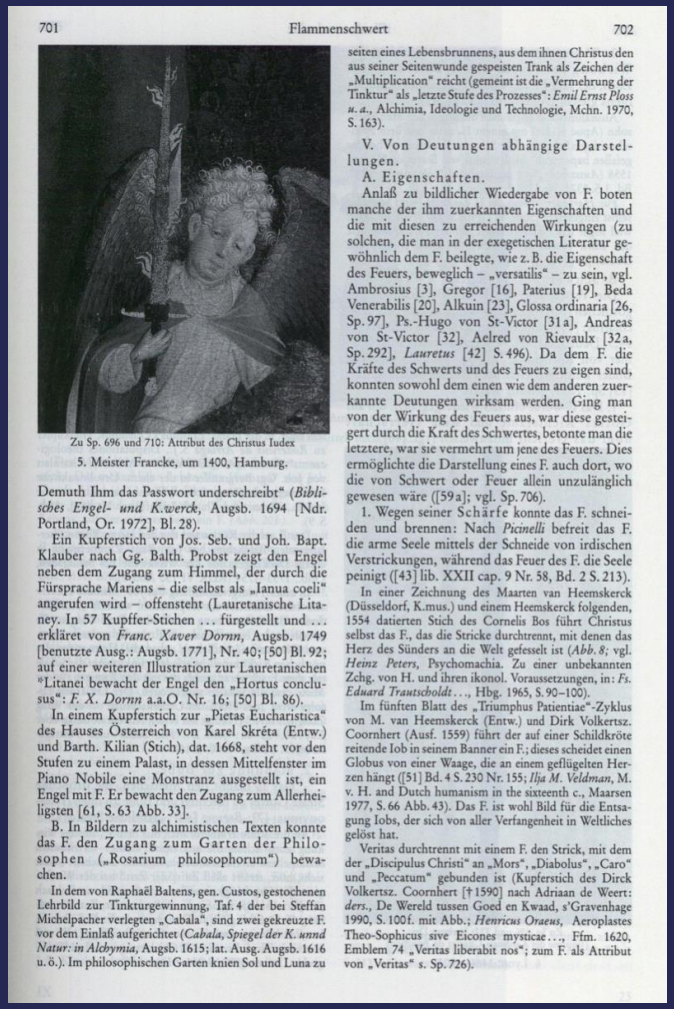
Automatische [a] Extraktion [s] fachterminologischer [a] Mehrwortgruppen [s]

AS automatische extraktion

AS fachterminologische mehrwortgruppe

Reallexikon zur Deutschen Kunstgeschichte - RDK

- Nachschlagewerk zur Realienkunde seit 1937
- Hohe Zahl an kunsthistorischen Sachbegriffen und Eigennamen (Personennamen, Geografika), spezifische Fachinformationen



701 Flammenschwert 702

Zu Sp. 696 und 710: Attribut des Christus Iudex
5. Meister Francke, um 1400, Hamburg.

Demuth Ihm das Passwort underschreibt" (*Biblisches Engel- und K.werck*, Augsb. 1694 [Ndr. Portland, Or. 1972], Bl. 28).

Ein Kupferstich von Jos. Seb. und Joh. Bapt. Klauber nach Gg. Balth. Probst zeigt den Engel neben dem Zugang zum Himmel, der durch die Fürsprache Mariens – die selbst als „Ianua coeli" angerufen wird – offensteht (Lauretanische Litaney. In 57 Kupffer-Stichen ... fürgestellt und ... erkläret von *Franc. Xaver Dornn*, Augsb. 1749 [benutzte Ausg.: Augsb. 1771], Nr. 40; [50] Bl. 92; auf einer weiteren Illustration zur Lauretanischen *Litanei bewacht der Engel den „Hortus conclusus": *F. X. Dornn* a.a.O. Nr. 16; [50] Bl. 86).

In einem Kupferstich zur „Pietas Eucharistica" des Hauses Österreich von Karel Skréta (Entw.) und Barth. Kilian (Stich), dat. 1668, steht vor den Stufen zu einem Palast, in dessen Mittelfenster im Piano Nobile eine Monstranz ausgestellt ist, ein Engel mit F. Er bewacht den Zugang zum Allerheiligsten [61, S. 63 Abb. 33].

B. In Bildern zu alchimistischen Texten konnte das F. den Zugang zum Garten der Philosophen („Rosarium philosophorum") bewachen.

In dem von Raphaël Baltens, gen. Custos, gestochenen Lehrbild zur Tinkturgewinnung, Taf. 4 der bei Steffan Michelpacher verlegten „Cabala", sind zwei gekreuzte F. vor dem Einlaß aufgerichtet (*Cabala, Spiegel der K. unnd Natur: in Alchymia*, Augsb. 1615; lat. Ausg. Augsb. 1616 u. ö.). Im philosophischen Garten knien Sol und Luna zu seiten eines Lebensbrunnens, aus dem ihnen Christus den aus seiner Seitenwunde gespeisten Trank als Zeichen der „Multiplication" reicht (gemeint ist die „Vermehrung der Tinktur" als „letzte Stufe des Prozesses": *Emil Ernst Ploss u. a.*, Alchimia, Ideologie und Technologie, Mchn. 1970, S. 163).

V. Von Deutungen abhängige Darstellungen.

A. Eigenschaften.

Anlaß zu bildlicher Wiedergabe von F. boten manche der ihm zuerkannten Eigenschaften und die mit diesen zu erreichenden Wirkungen (zu solchen, die man in der exegetischen Literatur gewöhnlich dem F. beilegte, wie z. B. die Eigenschaft des Feuers, beweglich – „versatilis" – zu sein, vgl. Ambrosius [3], Gregor [16], Paterius [19], Beda Venerabilis [20], Alkuin [23], Glossa ordinaria [26, Sp. 97], Ps.-Hugo von St-Victor [31a], Andreas von St-Victor [32], Aelred von Rievaulx [32a, Sp. 292], *Lauretus* [42] S. 496). Da dem F. die Kräfte des Schwerts und des Feuers zu eigen sind, konnten sowohl dem einen wie dem anderen zuerkannte Deutungen wirksam werden. Ging man von der Wirkung des Feuers aus, war diese gesteigert durch die Kraft des Schwertes, betonte man die letztere, war sie vermehrt um jene des Feuers. Dies ermöglichte die Darstellung eines F. auch dort, wo die von Schwert oder Feuer allein unzulänglich gewesen wäre ([59a]; vgl. Sp. 706).

1. Wegen seiner Schärfe konnte das F. schneiden und brennen: Nach *Picinelli* befreit das F. die arme Seele mittels der Schneide von irdischen Verstrickungen, während das Feuer des F. die Seele peinigt ([43] lib. XXII cap. 9 Nr. 58, Bd. 2 S. 213).

In einer Zeichnung des Maarten van Heemskerck (Düsseldorf, K.mus.) und einem Heemskerck folgenden, 1554 datierten Stich des Cornelis Bos führt Christus selbst das F., das die Stricke durchtrennt, mit denen das Herz des Sünders an die Welt gefesselt ist (*Abb. 8*; vgl. *Heinz Peters*, Psychomachia. Zu einer unbekannten Zchg. von H. und ihren ikonol. Voraussetzungen, in: *Fs. Eduard Trautscholdt* ..., Hbg. 1965, S. 90–100).

Im fünften Blatt des „Triumphus Patientiae"-Zyklus von M. van Heemskerck (Entw.) und Dirk Volkertsz. Coornhert (Ausf. 1559) führt der auf einer Schildkröte reitende Iob in seinem Banner ein F.; dieses scheidet einen Globus von einer Waage, die an einem geflügelten Herzen hängt ([51] Bd. 4 S. 230 Nr. 155; *Ilja M. Veldman*, M. v. H. and Dutch humanism in the sixteenth c., Maarsen 1977, S. 66 Abb. 43). Das F. ist wohl Bild für die Entsagung Iobs, der sich von aller Verfangenheit in Weltliches gelöst hat.

Veritas durchtrennt mit einem F. den Strick, mit dem der „Discipulus Christi" an „Mors", „Diabolus", „Caro" und „Peccatum" gebunden ist (Kupferstich des Dirck Volkertsz. Coornhert [†1590] nach Adriaan de Weert: *ders.*, De Wereld tussen Goed en Kwaad, s'Gravenhage 1990, S. 100f. mit Abb.; *Henricus Oraeus*, Aeroplastes Theo-Sophicus sive Eicones mysticae ..., Ffm. 1620, Emblem 74 „Veritas liberabit nos"; zum F. als Attribut von „Veritas" s. Sp. 726).

Quelle: Flammenschwert RDK, Spalten 701, 702, Band 9

Mehrwortgruppen im Reallexikon zur Deutschen Kunstgeschichte

- Kombinationen von Sachbegriffen
  „Darbringung im Tempel“
- Sachbegriffe und Personennamen
  „Verkündung an Maria“
- Sachbegriffe und Geografika
  „Flucht nach Ägypten“
- Funktionswörter (Präpositionen, Konjunktionen, Artikel)

Mehrwortgruppen mit Funktionswörtern im Reallexikon zur Deutschen Kunstgeschichte

- Syntaktisch abgeschlossene MWGs durch die Verbindung von Fachtermen und Funktionswörtern
  „Szene aus dem Leben Jesu“

- Inhaltliche Spezifizierung durch komplexere MWG-Konstruktionen
  Leben, Szene → „dargestellte Szene aus dem Leben Christi“

- Bedeutungsdifferenzierung einzelner Begriffe
  Zeit → „Kunst der altchristlichen Zeit“

Praktisches Vorgehen – Erstellung von Wörterbüchern

- Modifikation vorhandener Wörterbücher, die kunsthistorisches Fachvokabular enthalten (Sachbegriffe, Personennamen, Geografika)

- Erstellung eines Wörterbuchs mit Funktionswörtern (Artikel, Konjunktionen, Präpositionen)

- Allgemeines Rechtschreibwörterbuch

Praktisches Vorgehen – Vergabe neuer Wortklassen

Wortklasse	Bedeutung
E	Kunsthistorische Fachbegriffe
A	Adjektiv
K	Komposita
C	Präposition
R	Artikel
U	Konjunktion

Praktisches Vorgehen – Definition von Extraktionsmustern

- Bildung der Extraktionsmuster:
  - Erstellung von Kriterien zur Musterbildung
  - Folgen typischen Erscheinungsformen zusammengesetzter Terme im Deutschen
  - gezielte Platzierung von Funktionswörtern:
    - Artikel vor Substantiven (RE) oder Adjektiv-Substantiv Verbindungen (RAE)
    - Funktionswörter werden nur innerhalb eines Wortmusters platziert, beginnen und enden nie mit einem Funktionswort

ERAE *„Symbol des eucharistischen Christus“*

AERAE *„berühmte Komposition der italienischen Malerei“*

Ergebnisse der Mehrwortgruppenextraktion

- Erstellung von Kriterien zur Bewertung fachterminologischer Mehrwortgruppen zur qualitativen Beurteilung der Extraktionsmuster und daraus resultierenden Mehrwortgruppen:
  - Semantische Abgeschlossenheit
  - Spezifizierung von Personennamen und Geografika
- Indexierung von ca. 8.000 doppelspaltigen Lexikonseiten
- Vier- und fünfteilige Erkennungsmuster liefern beste Ergebnisse

Substantiv-Funktionswort-Adjektiv-Substantiv

ERKE *„darstellung der himmelserscheinung maria“*

Adjektiv-Substantiv-Funktionswort-Substantiv

AKUE *„weißes leinentuch als altarbekleidung“*

Adjektiv-Substantiv-Funktionswort-Adjektiv-Substantiv

AERAE *„epische dichtung der höfischen zeit“*

Wortmuster, die Personennamen oder Geografika spezifizieren

Personennamen:
ECEE *„altar von georg raphael”*

Geografika:
KCREE *„wandgemälde in der capella greca“*

Fazit – Erfolgreiche Extraktion von Mehrwort-gruppen mit Hilfe algorithmischer Verfahren

- Fachterminologische Mehrwortgruppen mit kunsthistorischem Bezug werden extrahiert
- Syntaktisch komplex strukturierte Mehrwortkonstruktionen durch Funktionswörter im Extraktionsalgorithmus
- Analyse umfangreicher Dokumentkollektionen hinsichtlich enthaltener Mehrwortgruppen möglich
- Modifizierung der Ergebnisse durch Korrektur der Wörterbücher und/oder Wortklassen möglich
- Allerdings: Analyse sehr zeitintensiv, Fachwissen zur Kunstgeschichte notwendig

Kontakt

Juliane Bredack

julianebredack@web.de