

# Text Mining in historical text books

A new project on information behavior in the humanities



Hildesheim, 6. Oct. 2013

## Partner



- **Georg Eckert Institute for International Textbook Research.**  
**Member of the Leibniz Association**  
**in Braunschweig**



**DIPF**

Bildungsforschung und Bildungsinformation

- **German Institute for International Educational Research**  
**Member of the Leibniz Association**  
**in Frankfurt a.M.**

# Text Mining

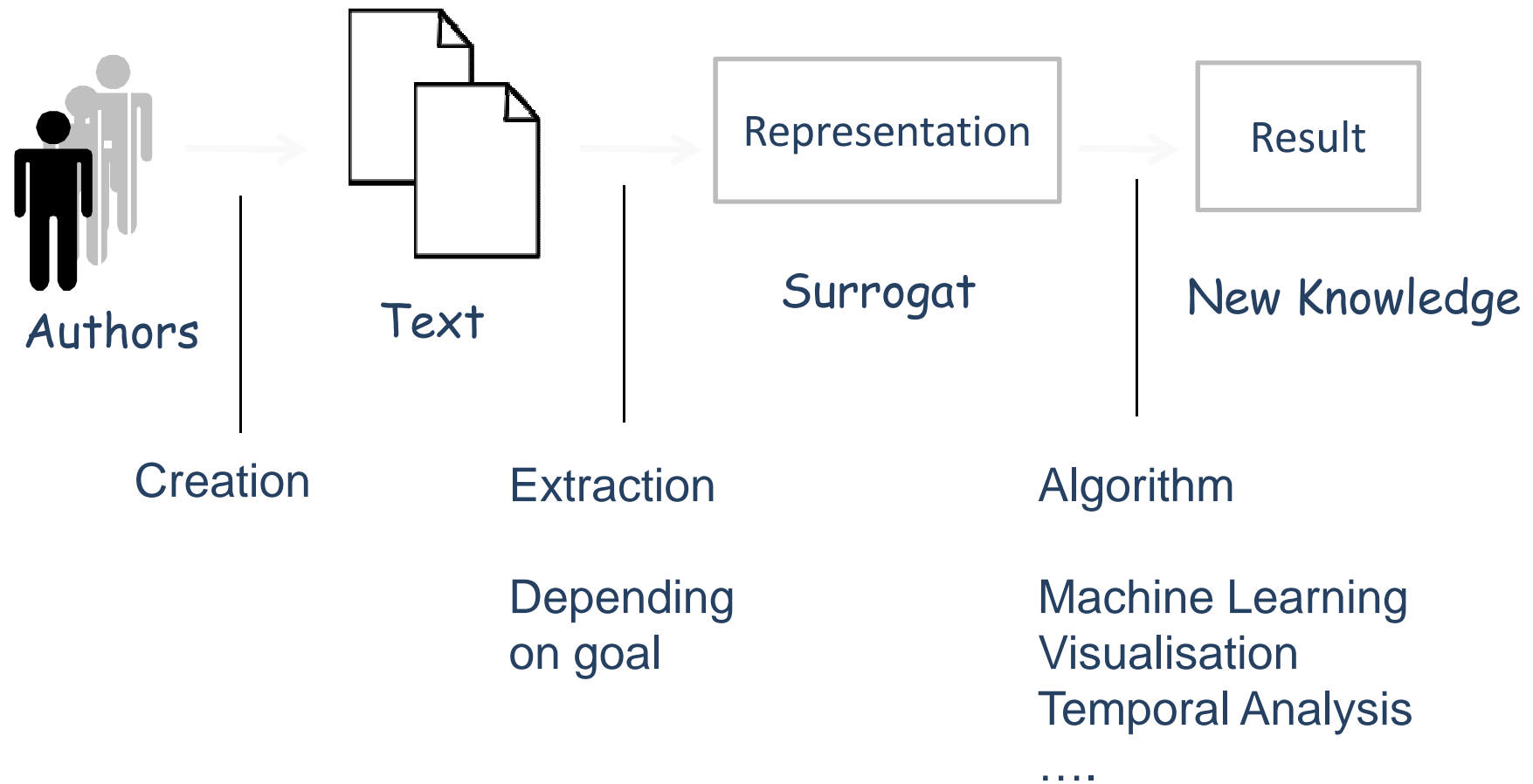
*Trends*

- “Discover useful and previously unknown “gems” of information in large text collections” (Avaquest 2002)
- „Text Mining is the extraction of knowledge out of many texts. .. The newly created knowledge could not be read from one single text but only by considering patterns over many texts ... „ (Mandl 2013 @ KSS)
- „Process of deriving high quality information from text” (Choi 2013 @ LWA)

*Patterns*

*Associations*

# Text Mining





# Die Welt der Kinder

# Die Welt der Kinder

The World of the Children

Knowledge about the world and meaning  
of world in childrens text books and  
children novels between 1850 und 1918



# Corpora

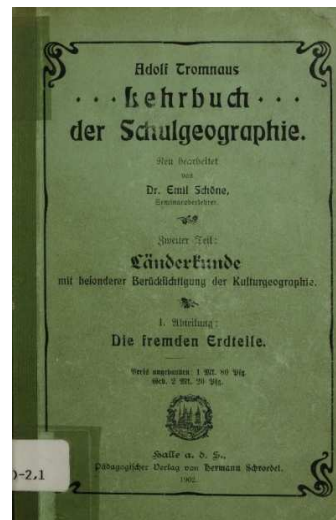
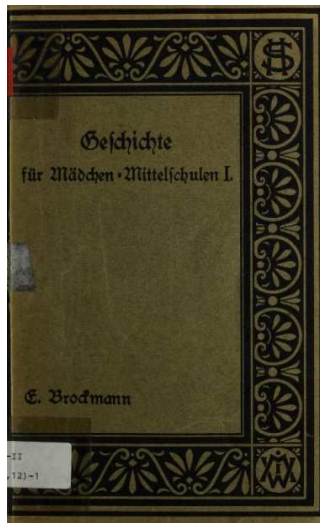
gei.digital

MDZ Münchener  
Digitalisierungszentrum  
Digitale Bibliothek



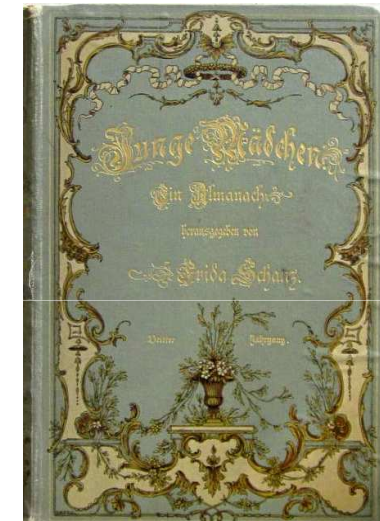
Universitätsbibliothek Braunschweig

## Text books (History and Geography)

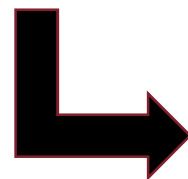


Knowledge from school books

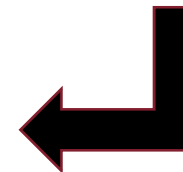
## Childrens' and teens' books



Popular Knowledge



Knowledge  
about the world



# Corpora

## ● GEI-Digital

- All German history text books of the „Kaiserreich“
- Some 1.800 books (350.000 pages)
- OCR processed
- Metadaten in METS/MODS-Structure
- Digitalisation of Geograpy books ongoing
- Until 2014 some 3.000 books more will be available (500.000 pages)
- TEI-Format



# Corpora

- Youth literature from UB Braunschweig
  - 2500 books,
  - Digitalisation ongoing
  - Genres: Adventure, Robinsonades, songbooks, moral and religious tales

# Text Books

- Text book in 19th century as a primary tool to access world knowledge
- Children were obliged to visit school
  - Between 1871 and 1918: 836 *new* history text books were published
  - Between 1890 and 1918: as many books were published as between 1700 and 1850

# *Areas of Transmitted Knowledge*

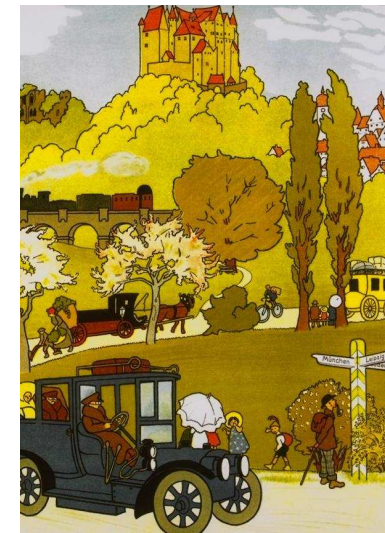
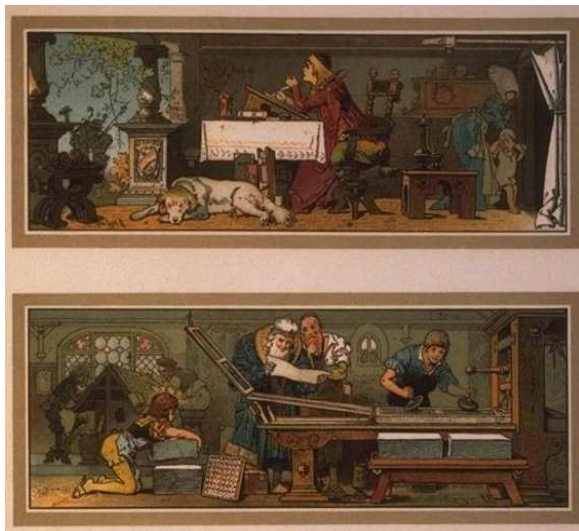
*„Foreign“ Continents,  
Countires and Peoples,  
Colonies, Discoveries,  
Expeditions,*

*War and Peace,  
political treaties,  
Regents and states  
men*

*Science and Culture,  
Urbanisation and  
Industrialisation,  
Pauperism, Social  
politics and family*

*Transport,  
Communication, Trade,  
Industry, Technology,  
environment and nature*

# Topics



Mandl: Text Mining in historical text books

# Goals

- Semantic, categorial and symbolic knowledge about the world
- How did the authorities interpret the world and transfer this knowledge to the children
- diachronic perspective
- Differences between parts of the corpora



# Topics

INHALT:		
CONTENTS:		CONTENU:
1. Entrance.	1. Einzug.	1. L'Arrivée.
2. The first white Man.	2. Der erste Weiße.	2. L'Apparition du Blanc.
3. A Professor of Zoology.	3. Ein Professor der Zoologie.	3. Un Professeur de Zoologie.
4. Black Love.	4. Schwarze Liebe.	4. Amour nègre.
5. Ostrich Parade.	5. Straußen-Parade.	5. Marche autrichienne.
6. Sunday Evening Walk.	6. Sonntagabend-Spaziergang.	6. Promenade de Dimanche.
7. Father and Son.	7. Der Vater mit dem Sohne.	7. Père et Fils.
8. The new Waistcoat.	8. Die neue Weste.	8. Le Gilet neuf.
9. The Ganymede of the Tropics.	9. Tropyen-Ganymed.	9. Ganymède dans les Tropes.
10. Flying Fish.	10. Fliegende Fische.	10. Les Poissons volants.
11. War-dance.	11. Kriegstanz.	11. Danse des Guerriers.
12. Gigantic Tarantula.	12. Riesen-Tarantel.	12. La Tarantule géante.

24 Drittes Kapitel.

Jahr hindurch und freisches Grün bedeckt schon Anfangs Februar die Berge, deren Schneehülle durch den warmen Hauch des Südwindes rasch abgestreift wird.

Helmuth von Rolffe hatte sich unterdessen häuslich eingerichtet. So wenig anmutig auch der erste Eindruck war, den er von Konstantinopel empfangen, so fühlte er sich doch nicht so unbehaglich in dem fremden Lande. Die Zuverlässigkeit und Güte seines Wirts trugen allerdings viel dazu bei, und ohne ihn würde er das Ungewohnte der Fremde wahrscheinlich mehr empfunden haben. Nachdem er sich von den Strapazen seiner Reise erholt, ersuchte er den Gesandten, ihn seinem zukünftigen Chef, dem Sersaskier, vorzustellen.

„Sie werden in ihm einen interessanten Mann kennen lernen“, sagte der gütige Landsmann, „jedenfalls ist er der nüchternste Mensch, den man sich denken kann. Biewohl er den Champagner nicht verschmäht, ist ihm ein Trunk aus den berühmten Quellen von Chamidje doch noch lieber. Er, der Reformfreund, hat den Türken in Sitte, Kleid und Denkweise gänzlich abgestreift. Von seiner Gewandtheit erhalten Sie einen ungefähren Begriff, wenn ich Ihnen mitteile, daß er sich während fünfundsiebzig Jahren in den höchsten Ämtern zu erhalten gewußt hat, was in der Türkei schon an ein Wunder streift. Kurzum, er ist nächst dem Sultan der mächtigste Mann im osmanischen Reiche.“

Das Buch.

1. Wer hat die Lieder nur erfunden  
In meinem lieben Bilderbuch?“ —  
Ein Dichter war's, der alle Wonnen  
Des lichten Frühlings in sich trug,  
Ein jugendfroher Sangesmeister,  
Der's lächelnd zu Papier gebracht,  
Was ihm die flinken Traumgesichter  
Verkündet in verschwiegener Nacht.

3. „Wie sind die Liedlein aber hier  
In mein geliebtes Buch gekommen?“ —  
Sein vielbeschriebenes Papier  
Hat er zum Drucker mitgenommen,  
Und dessen flinke Setzer haben  
In enger, dummer Werkstatt jetzt  
In schönen, klaren Druckbuchstaben  
Behende Vers auf Vers gesetzt. —

Die Landstraße

„Holla! Fröh! Anna! Hut auf! Rucksack her! Heute ist Malen- tag! Heute müssen wir wandern. Im hellen Sonnenschein wollen wir singen: Der Mai ist gekommen, die Bäume schlagen aus.“ So sagte die Mutter, und dann ging's hinaus aus der Stadt ins Feld. Puff, puff kommt die Eisenbahn daher und pufet: „Alles einsteigen!“ „Nein,“ sagt die Mutter, „du langer, ruhiger Wurm kannst ja nur unten durch die Berge kriechen. Wir wollen oben drüber klettern.“ „Huh, huh“ hustet die Lokomotive, spuckt Feuer und Rauch aus und ärgert sich, weil die drei nicht mitfahren wollen. Da rattert eine gelbe Postkutsche heran und fragt: „Nun, nicht mitreisen?“ Anna sagt aber: „Nein, dann können wir keine Blumen pflücken. Heute abend, wenn wir müde sind.“ — „Platz gemacht, ein Auto!“ Und das Auto fragt: „Wollt ihr nicht mit- fahren?“ Fröh fragt: „Kannst du uns wohl oben auf den Berg da tragen?“ „Nein, nein,“ sagt das Automobil. Da spricht Fröh: „Dann fahre nur zu, du Sauferwind, auf der staubigen Straße“



# Areas

INHALT:

CONTENTS:

CONTENU:

1. Entrance.	1. Einzug.	1. L'Arrivée.
2. The first white Man.	2. Der erste Weiße.	2. L'Apparition d'un Blanc.
3. A Professor of Zoology.	3. Ein Professor der Zoologie.	3. Un Professeur de Zoologie.
4. Black Love.	4. Schwarze Liebe.	4. Amour nègre.
5. Ostrich Parade.	5. Straußen-Parade.	5. Marche autruchienne.
6. Sunday Evening Walk.	6. Sonntagabend-Spaziergang.	6. Promenade de Dimanche.
7. Father and Son.	7. Der Vater mit dem Sohne.	7. Père et Fils.
8. The new Waistcoat.	8. Die neue Weste.	8. Le Gilet neuf.
9. The Ganyemed of the Tropics.	9. Tropen-Ganyemed.	9. Ganyémé dans les Tropes.
10. Flying Fish.	10. Fliegende Fische.	10. Les Poissons volants.
11. War-dance.	11. Kriegstanz.	11. Danse des Guerriers.
12. Gigantic Tarantula.	12. Riesen-Tarantel.	12. La Tarantule géante.

24 Drittes Kapitel.

Jahr hindurch und freies Grün bedeckt schon Anfangs Februar die Berge, deren Schneehülle durch den warmen Hauch des Südwindes rasch abgestreift wird.

Helmuth von **Maltz** hatte sich unterdessen häuslich eingerichtet. So wenig anmutig auch der erste Eindruck war, den er von **Kontantinescu** empfing, so fühlte er sich doch nicht so unbehaglich in dem fremden Lande. Die Zuverlässigkeit und Güte seines Wirts trugen allerdings viel dazu bei, und ohne ihn würde er das Ungeübte der **Reise** wahrscheinlich mehr empfunden haben. Nachdem er sich von den Strapazen seiner Reise erholt, ersuchte er den **Wirt** ihn seinem zukünftigen Chef, dem **Seraskier**, vorzustellen.

„Sie werden in ihm einen interessanten Mann kennen lernen“, sagte der gütige **Landmann**. „jedenfalls ist er der nüchternste Mensch, den man sich denken kann. Obwohl er den Champagner nicht verschmäht, ist ihm ein Trunk aus den berühmten Quellen von Chambridge doch noch lieber. Er, der Reformfreund, hat den **Wort** in Sinn, Kleid und Denkweise gänzlich abgestreift. Von seiner Gewandtheit erhalten Sie einen ungefähren Begriff, wenn ich Ihnen mitteile, daß er sich während fünfunddreißig Jahren in den höchsten Ämtern zu erhalten gewußt hat, was in der **Welt** schon an Wunder streift. Kurzum, er ist nächst dem **Emir** der mächtigste Mann im **osmanischen Reich**.“

Das **Buch**

„Wer hat die Lieder nur erfunden  
In meinem lieben Bilderbuch?“  
Ein Dichter war's, der alle Momen  
Des lichten Frühlings in sich trug,  
Ein jugendfroher Sangsmeister,  
Der's lächelnd zu **Papier** gebracht,  
Was ihm die flinken Traumgesichter  
Verkündet in verschwiegener Nacht.

„Wie sind die Liedlein aber hier  
In mein geliebtes Buch gekommen?“  
Sein vielbeschriebenes **Papier**  
Hat er zum **Druck** mitgenommen,  
Und dessen flinke Setzer haben  
In enger, dummer **Welt** jetzt  
In schönen, klaren Druckbuchstaben  
Behende Vers auf Vers gesetzt. —

Die **Landstraße**

„Holla! Fröh! Anna! Hut auf! Rucksack her! Heute ist Malen-  
tag! Heute müssen wir wandern. Im hellen Sonnenscheine wollen  
wir singen: Der Mai ist gekommen, die Bäume schlagen aus.“  
So sagte die Mutter, und dann ging's hinaus aus der Stadt ins  
Feld. Puff, puff kommt die **Landstraße** daher und pufet: „Alles  
einstiegen!“ „Nein,“ sagt die Mutter, „du langer, ruhiger Wurm  
kannst ja nur unten durch die Berge kriechen. Wir wollen oben  
drüber klettern.“ „Huh, huh“ hustet die **Landstraße**, spuckt Feuer  
und **Wand** aus und ärgert sich, weil die drei nicht mitfahren  
wollen. Da rattert eine gelbe **Postkutsche** heran und fragt: „Nun,  
nicht mitfahren?“ Anna sagt aber: „Nein, dann können wir keine  
Blumen pflücken. Heute abend, wenn wir müde sind.“ — „Platz  
gemacht, ein **Auto**!“ Und das Auto fragt: „Wollt ihr nicht mit-  
fahren?“ Fröh fragt: „Kannst du uns wohl oben auf den Berg  
da tragen?“ „Nein, nein,“ sagt das **Auto**. Da spricht Fröh:  
„Dann fahre nur zu, du Sauferwind, auf der staubigen **Straße**!“

# Untersuchungsfragen

Weltwissen, Weltbilder, Weltdeutung

Kontext Relevanz Semantik Transfer  
Deutungsmuster narrative Struktur Veränderung  
Diskurse kulturelle Übersetzungen  
Dispositive Repräsentation Zirkulation  
Kontinuität Transformation  
Selektion

Perspektiven und Vergleiche:

synchron

diachron

intermedial

transmedial

Selbst- und Fremdbilder, Fortschritts-, Krisen- und Bewältigungsdiskurse, ...

Rhein nach dem Elsaß vor. Nordgallien aber litt schwer unter den Raubzügen der Franken, und um Britanniens Küsten schwärmten die Hünen. Ohne angelsächsischer Seefahrer. Die Zeit der Wanderzüge

Selbst im vi  
fest verwachse  
große Beweg  
sondern auch  
Diese etwa 3  
wird gewöhnl

<mods:classification authority="GDZ">Kaiserreich Geschichtsschulbuec

- <mods:recordInfo>

<mods:recordIdentifier source="gbv-ppn">PPN665521057</mods:recordIdentifier>

</mods:recordInfo>

- <mods:titleInfo>

- <mods:title>

Lehrbuch für den Geschichts-Unterricht an Oberlyzeen und Studienans

</mods:title>

### III. Die Völkertwanderung.

Bereits um die Mitte des zweiten Jahrhunderts begann der Andrang ostgermanischer Völkerschwärme gegen die Donaulinie. Vandalen, Langobarden, Markomannen überschwemmten Jahrzehnte lang die römischen Grenzgebiete und drangen gelegentlich bis nach Oberitalien vor, ohne daß die Grenzlegionen das Weltreich ausreichend zu schützen vermochten. Dem Kaiser blieb nichts anderes übrig, als die Barbaren, die Land und Weideplätze beehrten, als Untertanen auf römischem Boden an der unteren Donau anzusiedeln. Dafür übernahmen sie die Pflicht des Grenzschutzes. Angesichts der germanischen Tapferkeit und Treue schien damit die Sicherheit der Reichsgrenze aufs neue verbürgt. Tatsächlich bildeten auch die neuen Kolonisten auf lange Zeit ein starkes Bollwerk, bis um die Mitte des dritten Jahrhunderts neue ostgermanische Haufen auftauchten, die Goten. Während im Westen die Allemen den Limes durchbrachen, fluteten die Goten von der Weichsel gegen das Schwarze Meer heran, überschritten die untere Donau und verheerten die römischen Grenzlande. Auf schnellen Booten suchten sie auch die Küsten des Ägäischen Meeres heim, plünderten zahlreiche Griechenstädte und zerstörten wertvolle



# *Methods, Tools, Technologies*

*Hermeneutics (intellectual text analysis) and qualitative Content Analysis*

and

*Tools for digital Analysis*



- *Opinion Mining* and *Topic Extraction* from historic text corpora
- Analysis and extraction of opinionated text parts
- Temporal Development of Topics and Opinions
- Adaptation of Tools to the specific demands of the historians and the structure of the Texts („domain specific text mining“)
- User oriented creation, evaluation and refinement of Software tools

# Tasks and Work Packages

- GEI

- Historical Research

- DIPF

- Software development

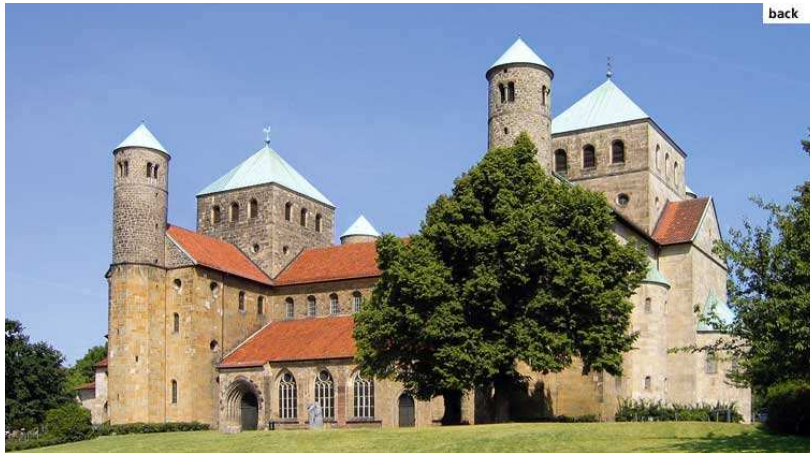
- U Hildesheim

- Information Behavior Analysis
- Requirements Specification
- Evaluation



**Wilson's process model of Ellis' behavioral framework  
of information seeking (adapted from Wilson 1999)**





## KONVENS 2014: Conference on Natural Language Processing

Uni Hildesheim  
Oktober 2014

Thank you for  
your attention