

Hypertextualisierung mit Topic Maps

Ein Ansatz zur Unterstützung des Textverständnisses bei der
selektiven Rezeption von Fachtexten

Eva Anna Lenz
Michael Beißwenger
Angelika Storrer
Universität Dortmund

Ein Textrezipient

- sucht in großen Dokumentenmengen
→ benötigt schnellen Überblick über Inhalte
- muss selektiv lesen
- ist ein Semi-Experte der Domäne

z.B. im Kontext interdisziplinärer Forschung, Studium und Ausbildung, Journalismus, Fachlexikographie.

Fokus des Vortrags: terminologiebedingte Probleme

- Wurde ein Fachterminus vom Autor eingeführt?
- nicht alle Termini bekannt / erkennbar

Mögliche Lösung:

- beim Lesen werden genau die notwendigen Wissensvoraussetzungen angeboten
 - hier: durch Links in einem Hypertext
- Linking nach Wissensvoraussetzungen

Bei terminologiebedingten Wissensvoraussetzungen:

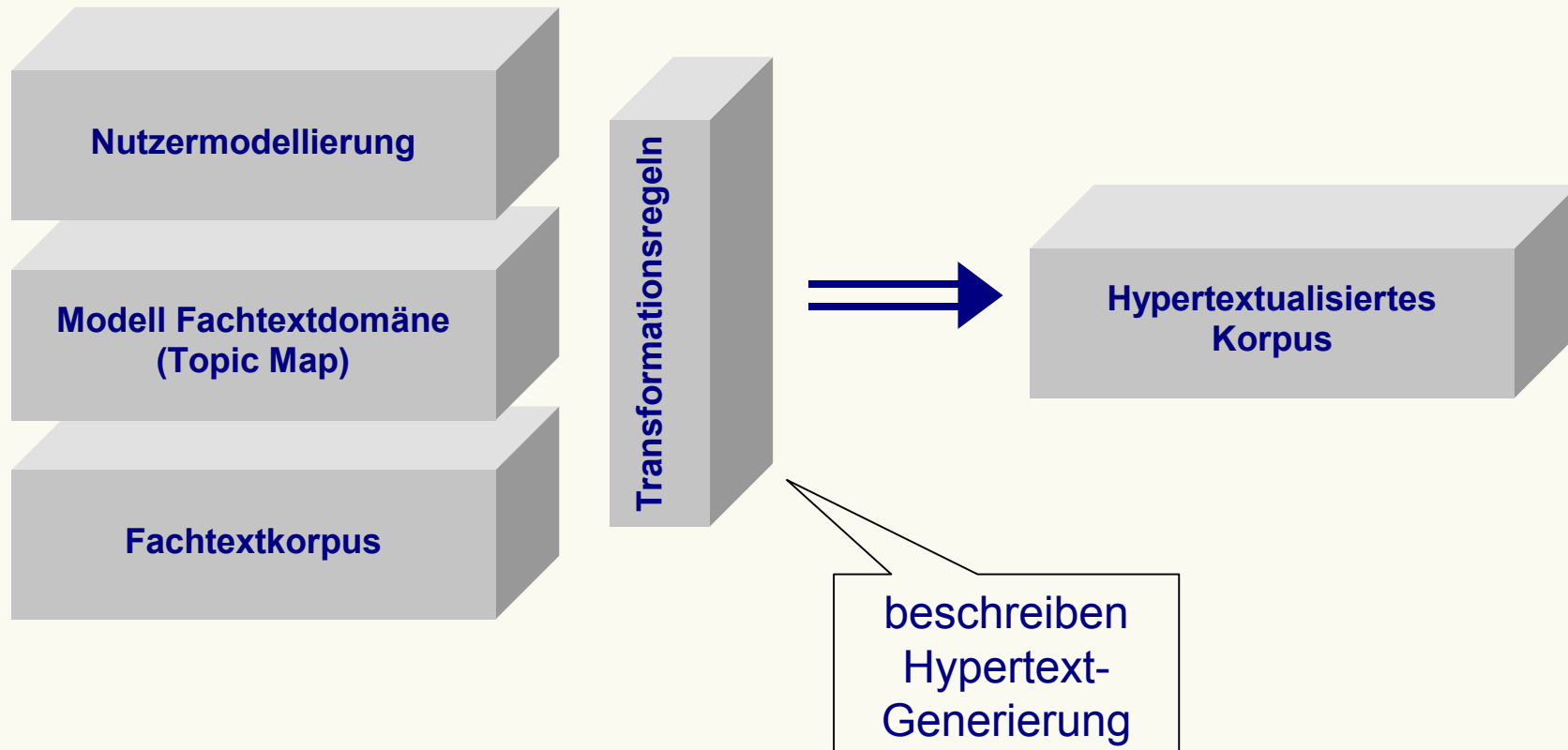
- *Kennzeichnung* eines Terminus als solcher
- *Informationen* über den Terminus
z.B. Links zu Definitionen bei einschlägigen Autoren
- *Darstellung der Beziehungen* des Terminus zu anderen Termini

- **HyTex**: Projektrahmen, Ziel, Architektur
- Modellierung terminologischen Wissens mit WordNet
- Implementierung durch Topic Maps
- Nutzung terminologischen Wissens zum Linking

Projektrahmen, Ziel, Architektur

- **HyTex**: Hypertextualisierung auf textgrammatischer Grundlage
- gefördert seit November 2001 von der DFG im Rahmen der Forschergruppe „Texttechnologische Informationsmodellierung“
- interdisziplinäres Projekt mit linguistischem Schwerpunkt
- Kontext des Semantic Web
 - Nutzung und Erprobung von Techniken (XML, Topic Maps, XSLT)
 - Ziel: mit konzeptionellen Ansätzen das Semantic Web für die menschliche Informationsverarbeitung besser erschließbar zu machen
- zunächst abgeschlossenes Textkorpus

- Ziel: Erforschung von konzeptionellen Strategien zur Hypertextualisierung eines Fachtextkorpus
- Domäne: Texttechnologie und Hypermedia
- Hypertextualisierungsstrategien basierend auf
 - textgrammatischem Markup
u.a. Koreferenz-Relationen, thematische Verweise, Definitionen und Gebrauch von Fachausdrücken
 - domänenspezifischem terminologischem Wissen
(Schwerpunkt des Vortrags)
- Ergänzung zu statistischen und KI-basierten Methoden
- Hypertext-Sichten: Erhaltung der ursprünglichen linearen Struktur

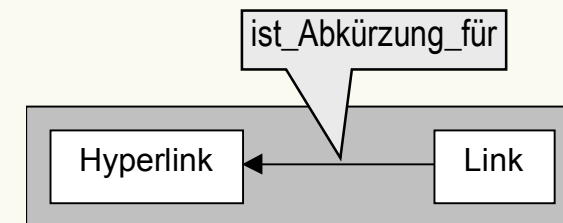
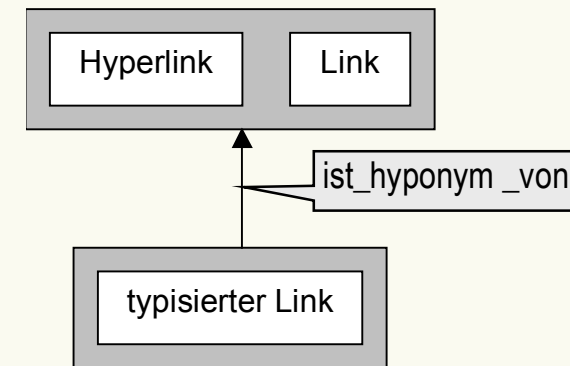
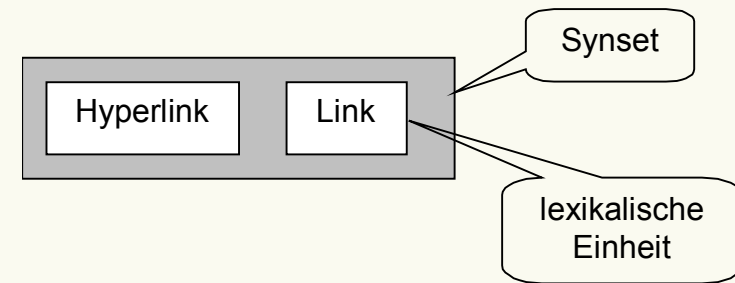


- alle Komponenten: XML
- Transformationsregeln werden automatisch in XSLT-Code übersetzt

Modellierung terminologischer Wissens mit WordNet

- lexikalische Datenbank
- ursprünglich für englische Sprache (Version 1.6: 138,838 Wörter),
- inzwischen auch für viele andere Sprachen (u.a. EuroWordNet)
- verschiedene Anwendungen im Bereich Language Engineering und Information Retrieval
- ursprünglich Allgemeinsprache, nicht Fachsprache
- nicht als Ontologie konzipiert, inzwischen aber in einigen Bereichen als Ontologie genutzt (mit Anpassungen)

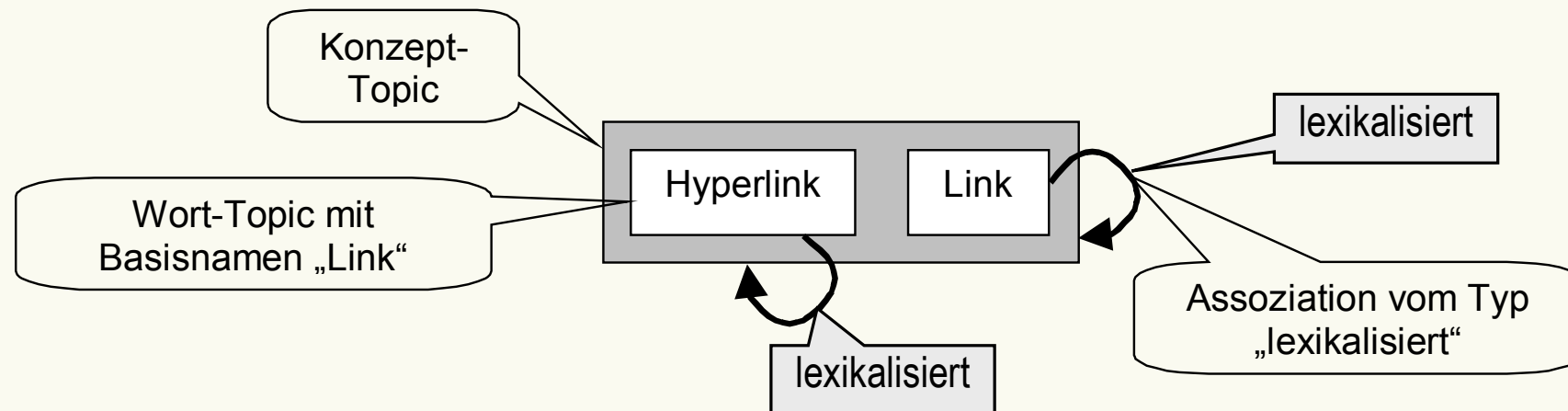
- Grundeinheit: Konzept (Begriff), repräsentiert durch ein „Synset“
- Synset besteht aus lexikalischen Einheiten (Wörtern), die das Konzept lexikalisieren
- semantisch-konzeptionelle Relationen zwischen Synsets, z.B. Hyponymie, Meronymie
- lexikalische Relationen zwischen lexikalischen Einheiten



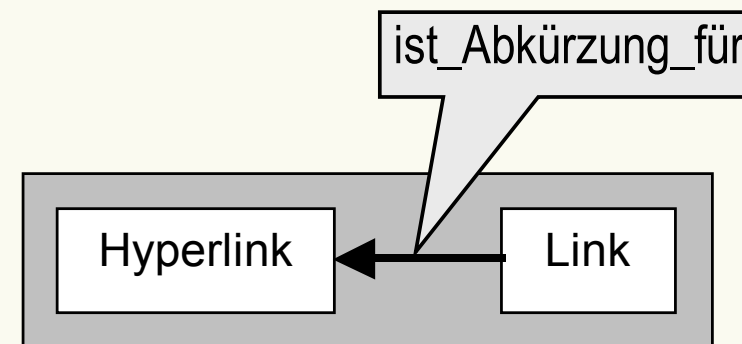
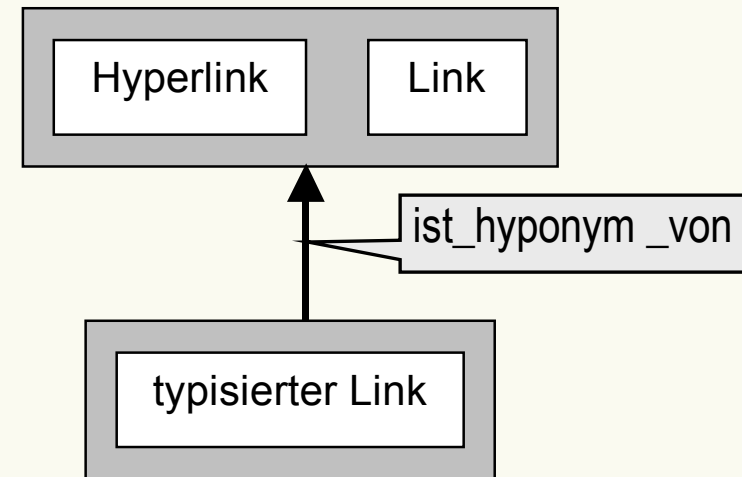
- wir nutzen nicht die Datenbank als solche, sondern übernehmen Organisation von WordNet für ein eigenes „TermNet“ der Domäne
- Erweiterungen
 - in Richtung Ontologie (z.B. class-instance)
XTM ist ein *XML-Standard*
 - neue domänenspezifische und lexikalische Relationen:
XTM ist eine Abkürzung für *XML Topic Maps*
(Akronymrelation)
Auszeichnungssprache ist der deutsche Ausdruck für *Markup Language*
(sprachkontaktbedingte Lexemkonkurrenz)
 - Klassifizierung der Termini, u.a. nach Varietäten:
Fachhochsprache (annotieren) vs.
Werkstattsprache (taggen)

Implementierung durch Topic Maps

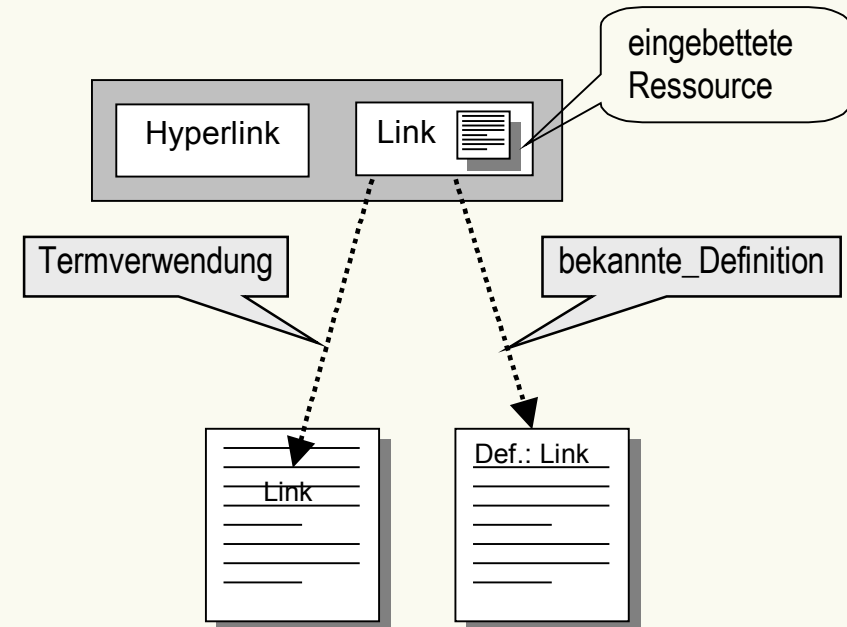
- für jeden terminologisierten Ausdruck (lexikalische Einheit) ein Topic mit dem Ausdruck als Basisnamen: Wort-Topic
- für jedes Synset ein Topic: Konzept-Topic
- jedes Wort-Topic wird mit zugehörigem Konzept-Topic durch eine Assoziation vom Typ `lexikalisiert` verbunden



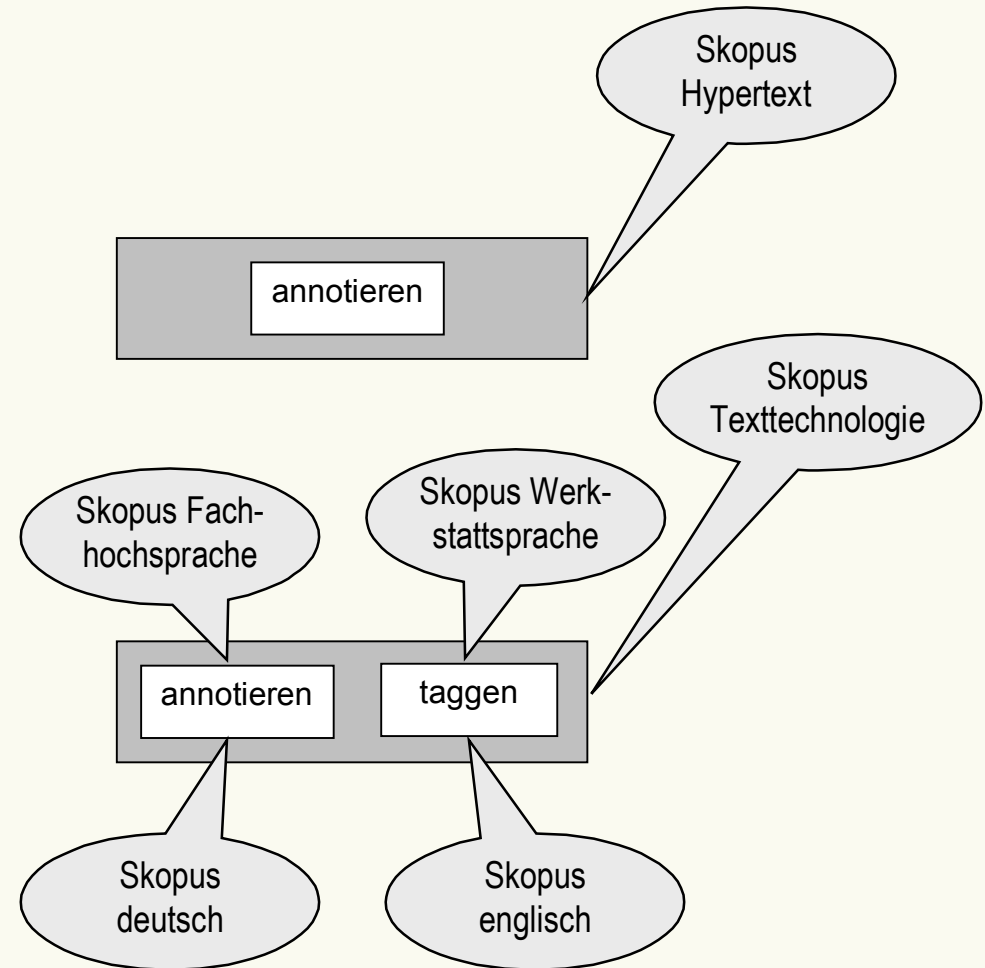
- für jede semantische Relation eine Assoziation des entsprechenden Typs
- dieser Typ ist wiederum ein Topic vom Typ `semantische_Relation`
- für jede lexikalische Relation eine Assoziation des entsprechenden Typs
- dieser Typ ist wiederum ein Topic vom Typ `lexikalische_Relation`



- Verbindungen von Wort-Topics zu Ressourcen durch Topic-Anker (occurrences):
 - zu jedem Vorkommen des Terminus
 - zu Erklärungen, Definitionen usw.
- eingebettete Ressourcen enthalten eine allgemeine Definition



- thematische Skopi auf der Ebene der Synsets:
Hypertext,
Texttechnologie, WordNet
- Skopi zur Unterscheidung von Sprachen:
deutsch, englisch
- Skopi zur Unterscheidung zwischen Varietäten:
Fachhochsprache,
Werkstattsprache



Nutzung terminologischen Wissens zum Linking

...nzepts von Tochtermann
...terscheidung von globalen
...tufte Definition von
...nstufen Modul – Kompo-
...scheidet er nach der
...zahl der in einem Link ©
...in denen ein Ausgangs-
...er verknüpft ist; 1:n-Links,
...it mehreren Zielankern
...denen mehrere Anker
...ngsrichtung miteinander zu
...rt sind. Im Linking-
...-Links vorgesehen; die
...zept des „Extended Link“
...n) sehen auch Links mit

LINK

Definitionen und Erklärungen

allgemeine Definition

Definition von Kühlen

FAQ

Verwendung in der KI

Verwandte Begriffe

andere Bezeichnung: Hyperlink

Gegenteil: –

Oberbegriffe: –

Unterbegriffe: Typisierter Link

ist Teil von: Hypertext

hat als Bestandteile: –

implementierbar durch: HTML-Linking-

Element

siehe auch: Anker

in der Community
bekannte Definition

Links zu verschie-
denen Textsorten

homographe Termini
anderer Fachbereiche

Synonym(e)

Antonym

Hyperonym

Hyponym(e)

Holonym

Meronym(e)

domänenspezifische
Relationen

Termini, die Wissens-
voraussetzung sind

- Filter
 - thematische Filter (Hypertext, Texttechnologie)
 - Varietätenfilter (Werkstattssprache, Fachhochsprache)
 - Sprachenfilter
- Generierung von Anmerkungen
 - z.B. taggen: Werkstattssprache für annotieren

...nzepts von Tochtermann
...terscheidung von globalen
...tufte Definition von
...nstufen Modul – Kompo-
...scheidet er nach der
...zahl der in einem Link ©
...in denen ein Ausgangs-
...er verknüpft ist; 1:n-Links,
...it mehreren Zielankern
...denen mehrere Anker
...ngsrichtung miteinander zu
...rt sind. Im Linking-
...-Links vorgesehen; die
...zept des „Extended Link“
...n) sehen auch Links mit

LINK

Definitionen und Erklärungen

allgemeine Definition

Definition von Kühlen

FAQ

Verwendung in der KI

Verwandte Begriffe

andere Bezeichnung: Hyperlink

Gegenteil: –

Oberbegriffe: –

Unterbegriffe: Typisierter Link

ist Teil von: Hypertext

hat als Bestandteile: –

implementierbar durch: HTML-Linking-Element

siehe auch: Anker

in der Community
bekannte Definition

Links zu verschie-
denen Textsorten

homographe Termini
anderer Fachbereiche

Synonym(e)

Antonym

Hyperonym

Hyponym(e)

Holonym

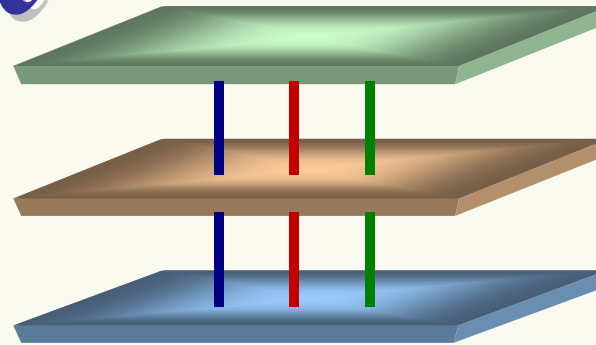
Meronym(e)

domänenspezifische
Relationen

Termini, die Wissens-
voraussetzung sind

HyTex

Hypertextualisierung
auf
textgrammatischer
Grundlage



<http://www.hytex.info/>