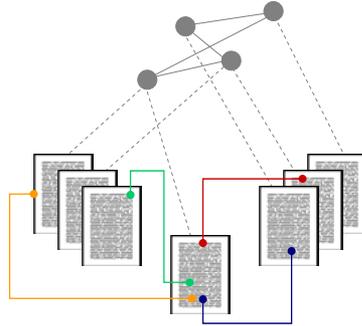


Terminologische Wortnetze für die Informationssuche in Fachtextcorpora



Michael Beißwenger | Angelika Storrer | Sandra Wellinghoff

■ Worum es geht

Vorstellung eines **Ansatzes zur Repräsentation von Fachwortschatz**,...

- der sich an den Modellierungskonstrukten von Wortnetzen orientiert;
- mit dem Ziel entwickelt wurde, Ähnlichkeiten und Unterschiede zwischen den Einheiten verschiedener Terminologien zum selben Gegenstand zu repräsentieren;
- gerade aufgrund seiner Orientierung an der Wortnetz-Modellierung von Wortnetzen geeignet erscheint, unterschiedliche Terminologien zu einem Gegenstand *trotz* der mit ihnen verbundenen unterschiedlichen Kategorisierungen in angemessener Weise zueinander in Beziehung zu setzen.

- 1) **Hintergrund: Fachsprachliches Terminologisieren = fachliches Kategorisieren (zu einem bestimmten Zweck)**
- 2) **Computerlesbare Repräsentation fachsprachlicher Terminologiesysteme: Modellierungsidee – Modellierungskonstrukte – Modellierungsebenen**
- 3) **Vorteile und Anwendungsfelder einer *wortnetzbasierten* Repräsentation gegenüber *term-* oder *konzeptbasierten* Ansätzen**
- 4) **Ausblick: Verknüpfung von *TermNet* mit *GermaNet***

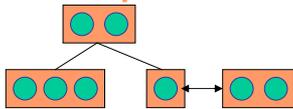
■ Hintergrund: *Term- und konzeptbasierte* Ansätze

termorientiert



Termorientierte Modellierungen terminologischer Thesauri (z.B. ISO 2788; ANSI/NISO Z3919) gehen davon aus, dass im Normalfall ein Terminus einem Konzept entspricht; wenn Benennungsalternativen vorliegen lässt sich eine davon als „preferred term“ bestimmen. Relationen wie „broader term“ (der Entsprechung der Hyperonymie) bestehen zwischen Termini.

konzeptbasiert



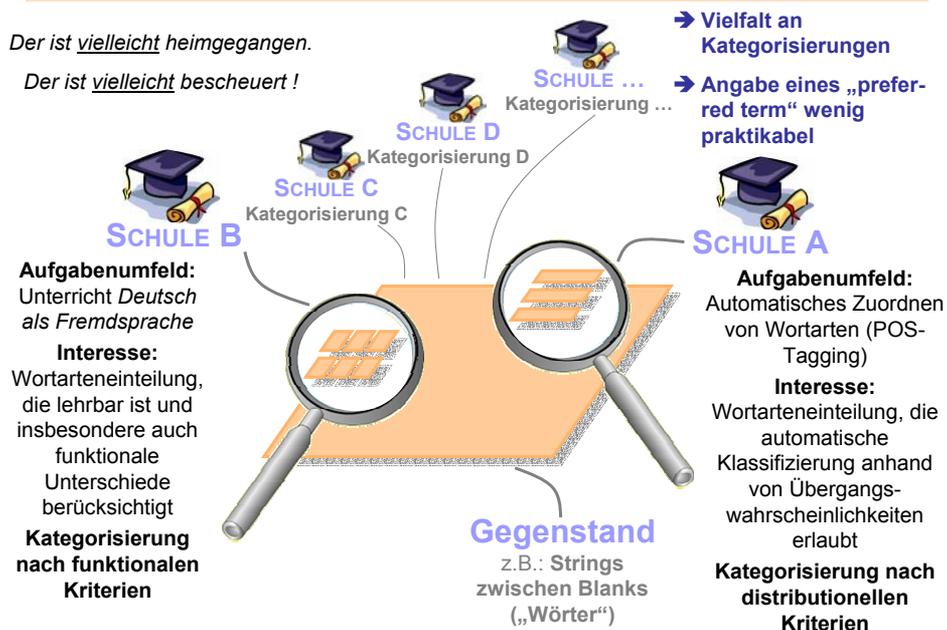
In **konzeptbasierten Ansätzen** werden Termini Konzepten zugeordnet; vertikale Relationen (Teil-Ganzes; Überordnung-Unterordnung) bestehen dann zwischen Konzepten. Auch hier kann zwar ein Terminus als „preferred term“ spezifiziert werden; eine direkte Relationierung zwischen Termini ist aber nicht vorgesehen.



■ Hintergrund: Kategorisieren von Weltausschnitten

Der ist vielleicht heimgegangen.

Der ist vielleicht bescheuert !



■ Hintergrund: Kategorisieren von Weltausschnitten

In diversen Einzelwissenschaften zu beobachten:

Ein Untersuchungsgegenstand –

mehrere (konkurrierende) terminologische Systeme

→ **Verschiedene theoretische Ansätze und Schulen** befassen sich mit demselben Gegenstand aus unterschiedlichen Perspektiven und mit verschiedenen Zielsetzungen und entwerfen **dafür zweckdienliche Beschreibungskategorien**

Die entsprechenden **Termini sind dann jeweils in einem dieser Systeme verankert** (durch ihren *valeur* bestimmt); ihre Bedeutung ist in mehr oder weniger formalen Definitionen (in Fachtexten) festgelegt

Zwischen den Konzepten, die die Termini dieser verschiedenen Systeme lexikalisieren, lassen sich allerdings Bezüge der **Kategorienähnlichkeit** herstellen. Diese kann unterschiedlich stark ausgeprägt sein.

Beispiele:

- (a) **Terminus X** aus **System 1** ist extensionsgleich mit **Terminus X** aus **System 2** (*DUDEN* „Verb“ vs. *KMK-Wortarten* „Verb“) **[IDEALFALL]**
- (b) **System 1** subsumiert unter eine Kategorie (**Terminus X**) weniger Einheiten als **System 2** (*Helbig/Buscha* „Adverb“ vs. *DUDEN* „Adverb“)
- (c) **System 1** subsumiert unter eine Kategorie (**Terminus X**) auch Einheiten, die in **System 2** einer anderen Kategorie (**Terminus Y**) zugerechnet werden (*DUDEN* „Adjektiv“ vs. *KMK-Wortarten* „Adjektiv“, „Numeralia“)



■ Hintergrund: Bezüge zwischen Kategorisierungssystemen

Unterscheidungen derart, wie wir sie diskutieren werden, wurden von vielen Philosophen gemacht, allerdings auf verschiedene Weise. Wie wir sehen werden, ist es jetzt üblich, eine zweifache Unterscheidung zwischen dem, was wir **Sinn**, und dem, was wir **Referenz** nennen werden, zu machen. Andere Termini, die für den gleichen oder zumindest einen ähnlichen Kontrast verwendet werden, sind: ‚Bedeutung‘ und ‚Referenz‘ (wobei ‚Bedeutung‘ eine engere Interpretation erhält als diejenige, die sie als alltäglicher vortheoretischer Ausdruck hat); ‚Konnotation‘ und ‚Denotation‘; ‚Intension‘ und ‚Extension‘.

John Lyons: *Semantik*. Bd. I. München 1980,
Kap. 7 „Referenz, Sinn, Denotation“ – 7.2. „Einleitende Bemerkungen“

Das Aufzeigen und das schnelle Auffinden von Bezügen nicht zwischen einzelnen Termini, sondern zwischen ganzen (terminologischen) Kategorisierungssystemen ist für viele Bereiche der Wissenschaftskommunikation von großer Bedeutung.



■ Unser Ansatz: Wortnetzbasierete Repräsentation

Wozu eine **wortnetzbasierete Modellierung** von Fachwortschatz?

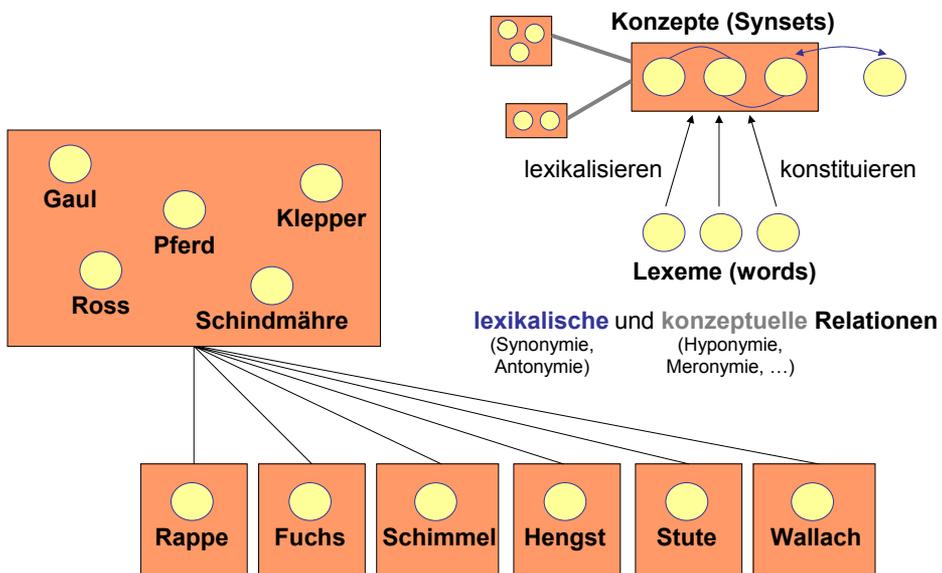
Ziel: Computerlesbare Repräsentation von Bezügen zwischen *verschiedenen* Terminologien in ein- und demselben Gegenstandsbereich



- **Projekt zur automatischen Hypertextualisierung** (Teilprojekt der verteilten DFG-Forschergruppe „Texttechnologische Informationsmodellierung“)
- Modellierung von (definierten) terminologischen Konzepten und Terminologien auf einem **Fachtextcorpus zur Domäne „Hypertextforschung und Texttechnologie“** (Umfang: 20 Dokumente / 113.666 lfd. Wortformen; darin 191 Definitionen)
- **TermNet** (205 Konzepte, 477 Lexeme): Bislang modelliert mit XML Topic Maps; geplant: Re-Modellierung in OWL-DL und Verknüpfung mit *GermaNet*

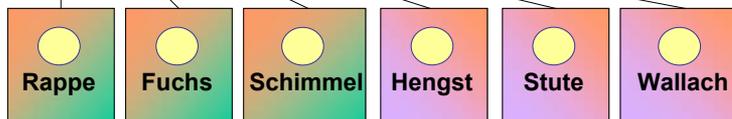
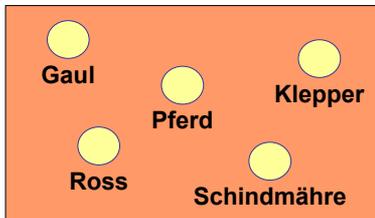


■ Modellierungskonstrukte in Wortnetzen



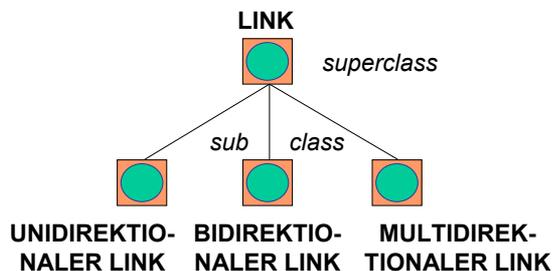
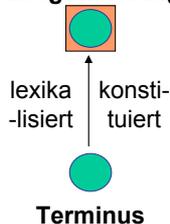
Charakteristika von Wortnetzen (u.a.)

- **Synsets** enthalten Wörter, die in mindestens einem Kontext füreinander austauschbar sind (also keine Äquivalenz / Extensionsgleichheit im strengen Sinne)
- **Hyperonymie** ist eine „kind-of“-Beziehung zwischen Konzepten; man kann sie als „weiche“ Form der strengen *superclass-subclass*-Beziehung reinterpretieren; sie ist aber explizit nicht in dieser strengen Form intendiert
- **Synsets** sind intern wenig strukturiert
- Die Modellierung von **extensional disjunkten Kohyponymen** (also von Subklassen auf derselben Hierarchieebene) ist nicht direkt möglich



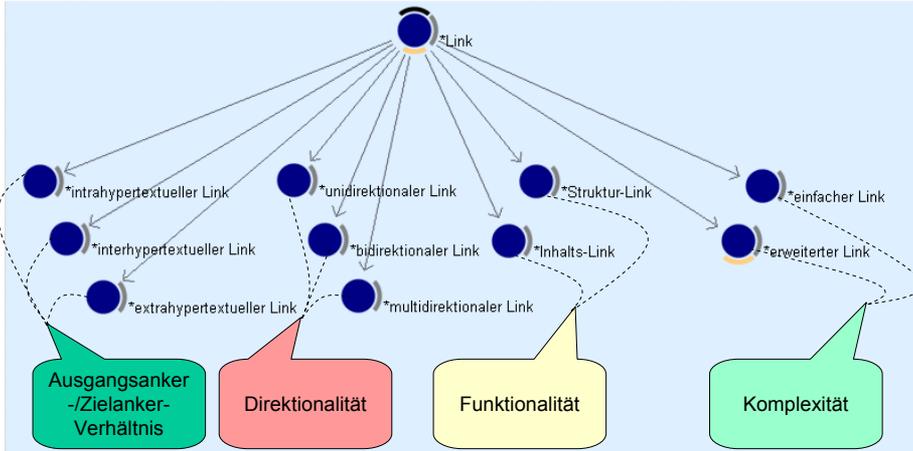
Übertragung des Wortnetzansatzes auf Fachwortschatz

Konzept / Kategorisierung



- Innerhalb eines terminologischen Systems entspricht ein Terminus meist genau *einem* Konzept.
- Vertikale Relationen wie Überordnungs- und Unterordnungsbeziehungen zwischen den Termini sind meist tatsächlich als Inklusion (also im Sinne von *subclass-superclass*) gedacht.
- Typisch für terminologische Kategorisierungen sind Kohyponymgruppen auf derselben Stufe, die jeweils durch ein Klassifikationsmerkmal motiviert und die wechselseitig disjunkt sind.

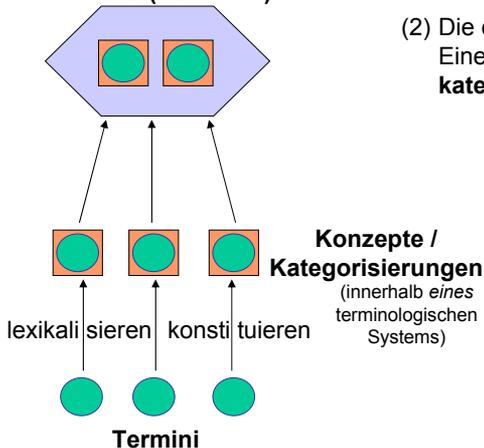
■ Übertragung des Wortnetzansatzes auf Fachwortschatz



- Typisch für terminologische Kategorisierungen sind Kohyponymgruppen auf derselben Stufe, die jeweils durch ein Klassifikationsmerkmal motiviert und die wechselseitig disjunkt sind.

■ Modellierungskonstrukte/-ebenen: *TermNet* – Wortnetze

Kategorienähnliche Konzepte 2. Stufe (*Termsets*)



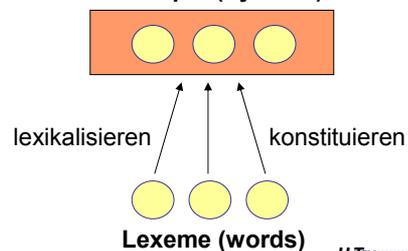
Termset:

- Die enthaltenen Einheiten sind **Konzepte**.
- Die enthaltenen Einheiten sind **kategorienähnlich**.

Synset:

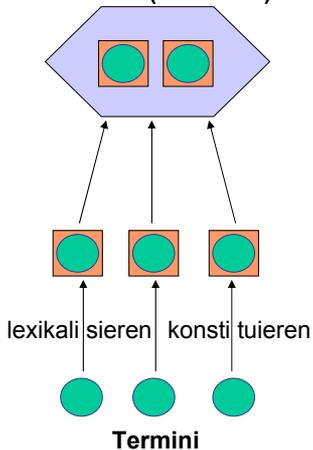
- Die enthaltenen Einheiten sind **Lexeme**.
- Die enthaltenen Einheiten sind **synonym**.

Konzepte (Synsets)



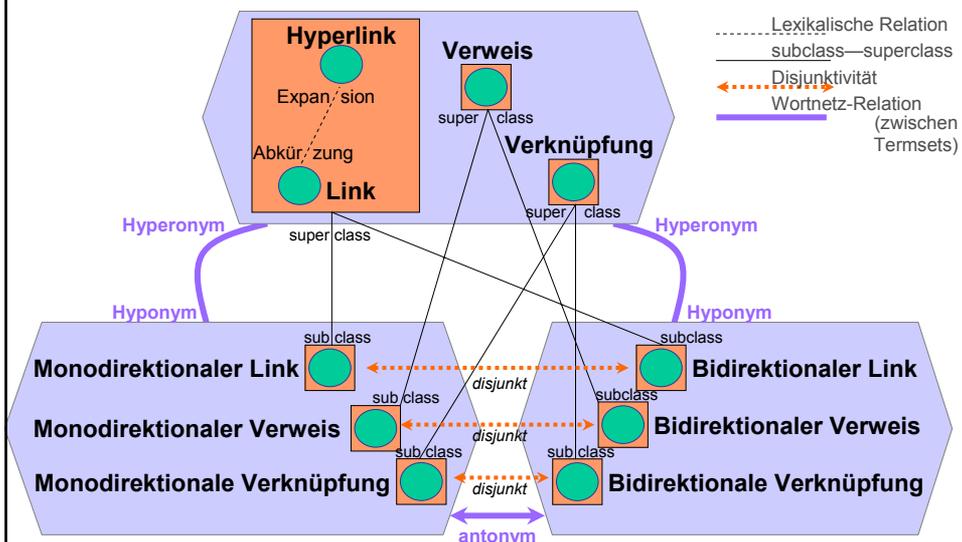
Modellierungskonstrukt: *Termset*

Kategorienähnliche Konzepte 2. Stufe (*Termsets*)



- **Termsets** enthalten Termini aus verschiedenen Terminologien, die extensional einen relativ großen Schnittbereich haben.
- Termsets sind durch „weiche“ **konzeptuelle Wortnetzrelationen** wie „Hyperonymie“ und „Meronymie“ miteinander verbunden. Wie bei allgemeinsprachlichen Wortnetzen auch, kann man für bestimmte Anwendungen diese Relationen „hart“ reinterpretieren.
- Die Einheiten von Termsets sind aber, im Gegensatz zu den Synset-Einheiten, in einem Text nicht gegeneinander austauschbar; der Autor bewegt sich ja gerade in einem bestimmten System.
- Bei den Relationen haben wir somit eine Differenzierung:
 - „harte“ **Relationen** bestehen zwischen den Einheiten desselben Systems;
 - „weiche“ **Relationen** bestehen zwischen Termsets.

Kategorienähnlichkeit: Wortnetzbasierte Repräsentation



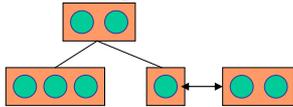
■ Kategorienähnlichkeit: Wortnetzbasierte Repräsentation

termorientiert



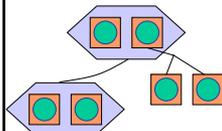
Termorientierte Modellierungen terminologischer Thesauri (z.B. ISO 2788; ANSI/NISO Z3919) gehen davon aus, dass im Normalfall ein Terminus einem Konzept entspricht; wenn Benennungsalternativen vorliegen lässt sich eine davon als „preferred term“ bestimmen. Relationen wie „broader term“ (der Entsprechung der Hyperonymie) bestehen zwischen Termini.

konzeptbasiert



In **konzeptbasierten Ansätzen** werden Termini Konzepten zugeordnet; vertikale Relationen (Teil-Ganzes; Überordnung-Unterordnung) bestehen dann zwischen Konzepten. Auch hier kann zwar ein Terminus als „preferred term“ spezifiziert werden; eine direkte Relationierung zwischen Termini ist aber nicht vorgesehen.

wortnetzbasiert



Mit einem **wortnetzbasierter Ansatz** wird es möglich, unterschiedliche terminologische Systeme in einer Domäne zueinander in Beziehung zu setzen (term- und konzeptorientierte Terminologiemodellierungen können hingegen jeweils nur *ein* System abbilden)



■ Anwendungen

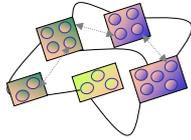
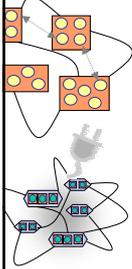
Information Retrieval:

Auswahl verschiedener Suchschärfen

Unterstützung der selektiven Hypertextrezeption (im Anwendungsfeld Wissenskommunikation):

Verlinkung von Fachtexten mit automatisch aus der Wortnetzrepräsentation erzeugten Glossarsichten

■ Vernetzung mit GermaNet



2 Philosophien:

- **Plug-in-Ansatz:** Termini werden unter ein allgemeinsprachliches Konzept gesetzt
- **Feature-Ansatz:** Verwendung von Domänen-Features, um terminologische Teilnetze zu kennzeichnen

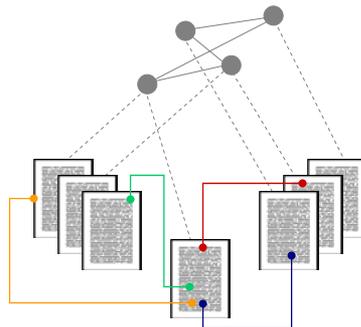
Ansatz (noch in Bearbeitung in Zusammenarbeit mit C. Kunze):
Plugin-Philosophie unter Nutzung von **OWL-properties**

Leitidee:

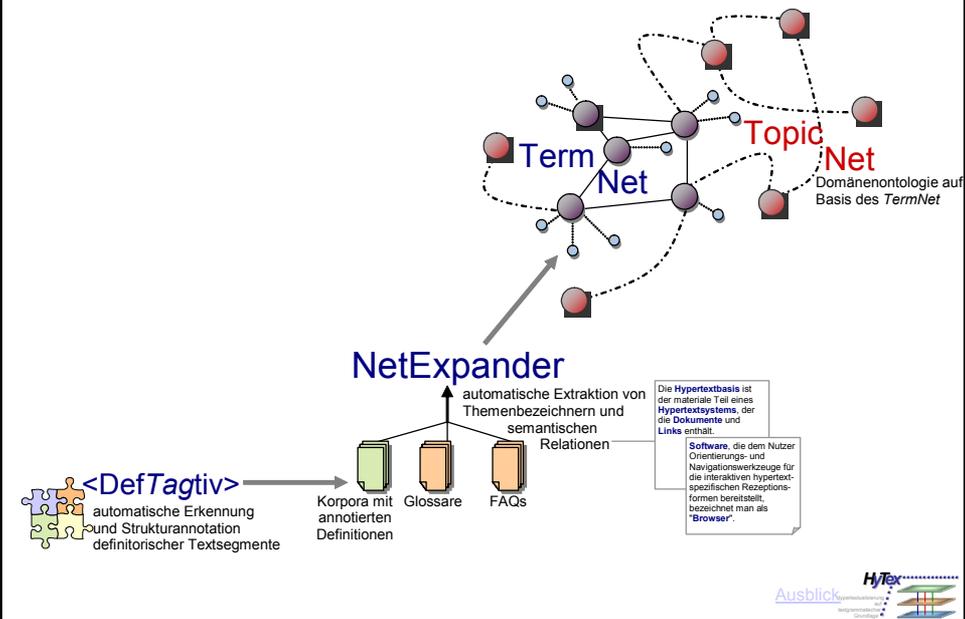
Vernetzung von **Termini mit Synsets**, in denen das sinnentsprechende Lexem zu finden ist (weil ja ein Terminus einem Konzept entspricht)



Terminologische Wortnetze für die Informationssuche in Fachtextcorpora



Ausblick: Ausbau von TermNet zum TopicNet



Ausblick: Topic Chaining auf Basis des TopicNet

