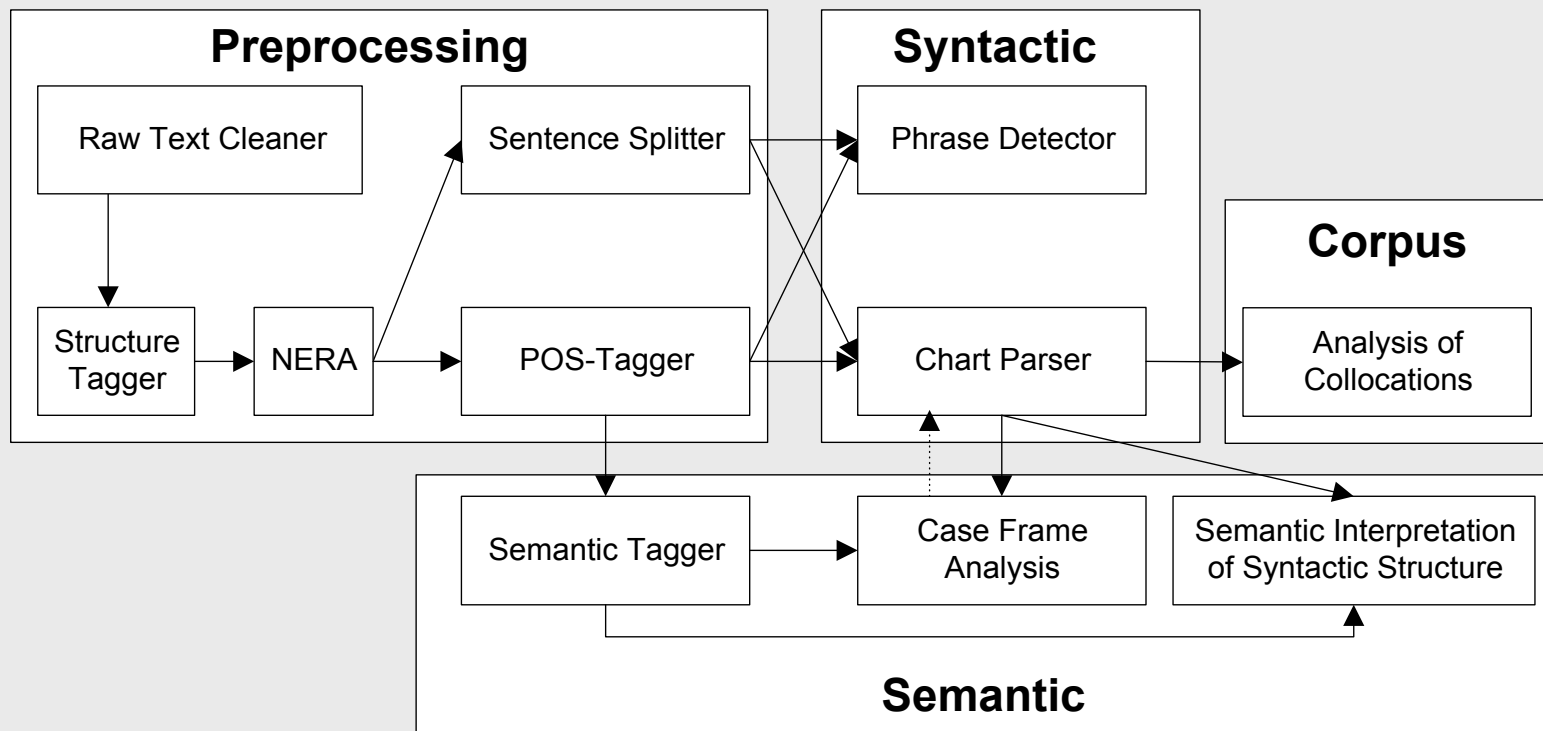


Processing Medline Abstracts: an Experiment in Resource Integration

Manuela Kunze, Chun Xiao, Dietmar Rösner

Otto-von-Guericke Universität Magdeburg
Fakultät für Informatik
Institut für Wissens- und Sprachverarbeitung

What is XDOC?

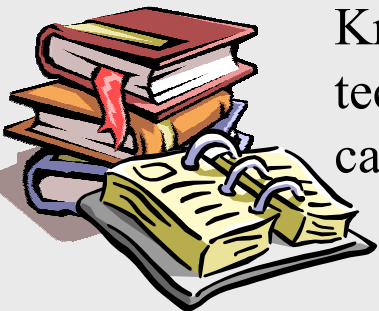


XDOC:

XML based *DOC*ument processing (for German)

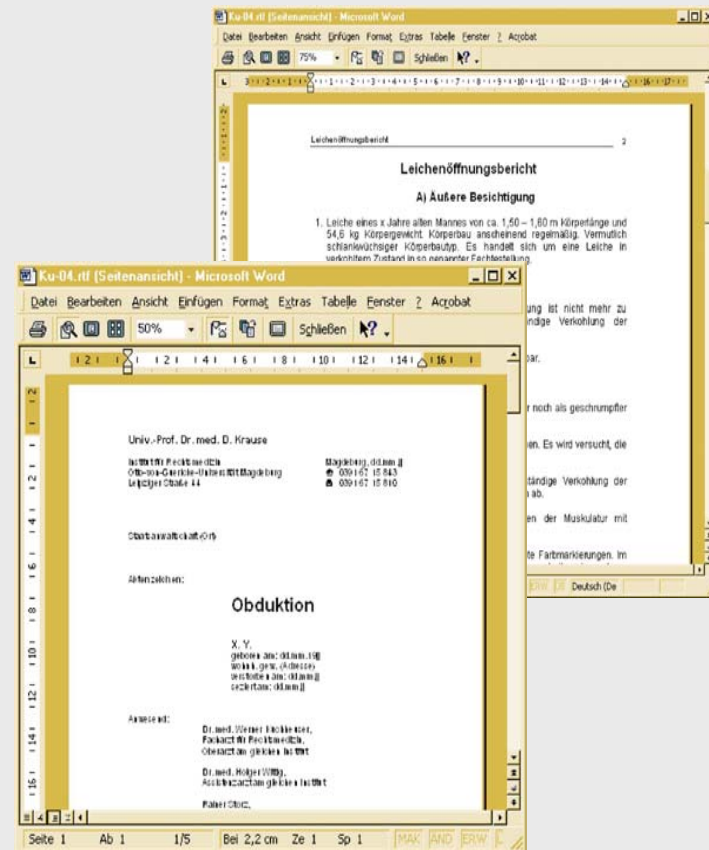
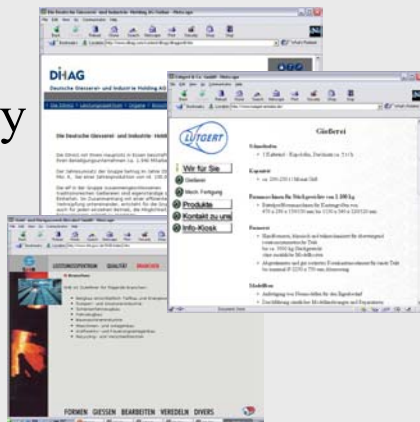
► inputs, results and resources: *XML* based

Corpus/Application Scenarios



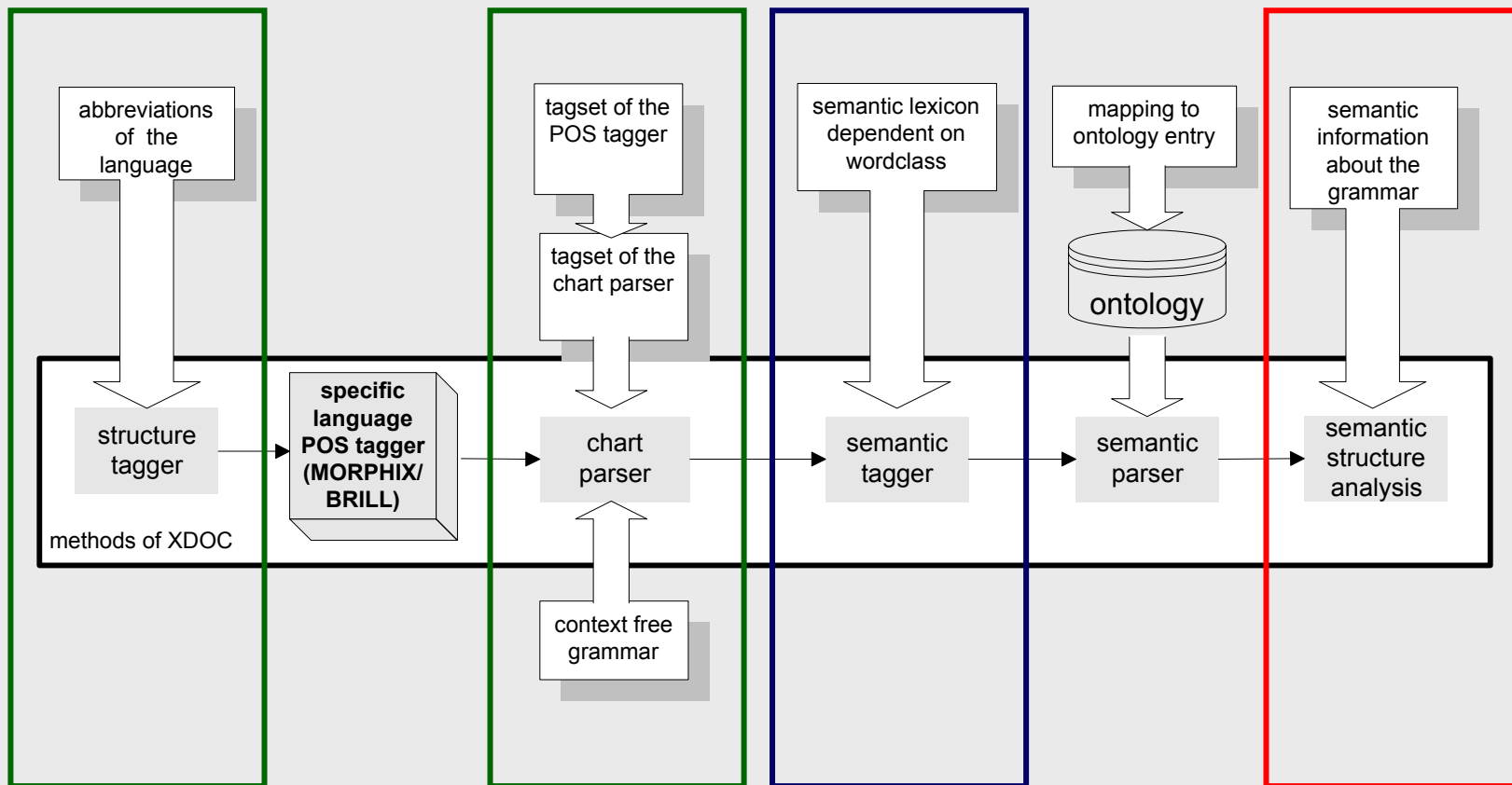
Knowledge acquisition from technical documentation about casting technology

Extraction of company and product profiles from webpages



Analysis of autopsy protocols for statistical summary about typical injuries at traffic accidents

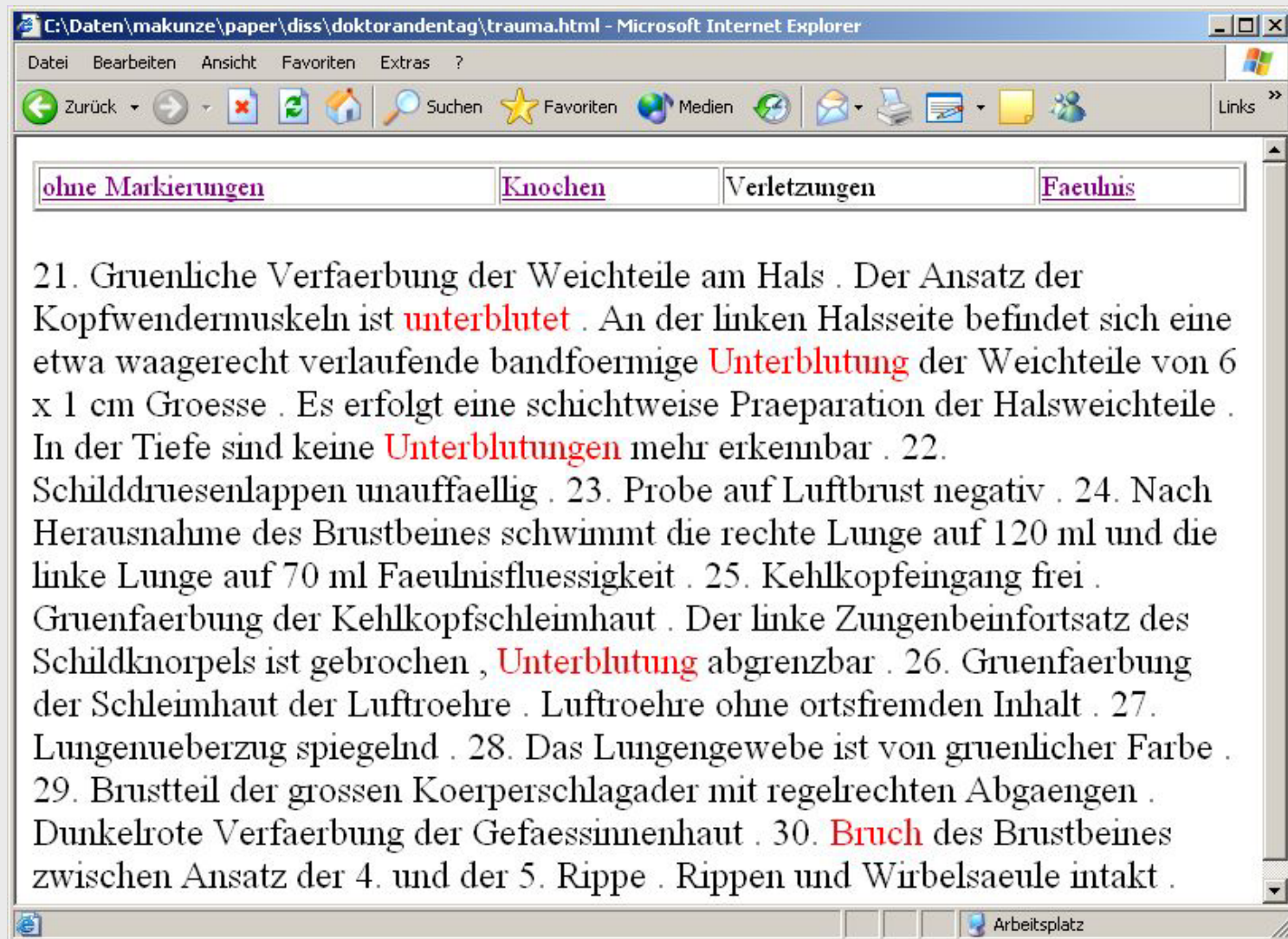
XDOC for English



Semantic Tagger

- classify content words into their semantic categories
- input data: a text tagged with POS tags
 - additional: annotated named entities
- resource: semantic lexicon
 - semantic interpretation of a token
 - information about case frame

Semantic Tagger



C:\Daten\makunze\paper\diss\doktorandentag\trauma.html - Microsoft Internet Explorer

Datei Bearbeiten Ansicht Favoriten Extras ?

Zurück Suchen Favoriten Medien Links

[ohne Markierungen](#) [Knochen](#) [Verletzungen](#) [Faeulnis](#)

21. Gruenliche Verfaerbung der Weichteile am Hals . Der Ansatz der Kopfwendermuskeln ist **unterblutet** . An der linken Halsseite befindet sich eine etwa waagrecht verlaufende bandfoermige **Unterblutung** der Weichteile von 6 x 1 cm Groesse . Es erfolgt eine schichtweise Praeparation der Halsweichteile . In der Tiefe sind keine **Unterblutungen** mehr erkennbar . 22. Schilddruesenlappen unauffaellig . 23. Probe auf Luftbrust negativ . 24. Nach Herausnahme des Brustbeines schwimmt die rechte Lunge auf 120 ml und die linke Lunge auf 70 ml Faeulnisfluessigkeit . 25. Kehlkopfeingang frei . Gruenfaerbung der Kehlkopfschleimhaut . Der linke Zungenbeinfortsatz des Schildknorpels ist gebrochen , **Unterblutung** abgrenzbar . 26. Gruenfaerbung der Schleimhaut der Luftroehre . Luftroehre ohne ortsfremden Inhalt . 27. Lungenueberzug spiegelnd . 28. Das Lungengewebe ist von gruenlicher Farbe . 29. Brustteil der grossen Koerperschlagader mit regelrechten Abgaengen . Dunkelrote Verfaerbung der Gefaessinnenhaut . 30. **Bruch** des Brustbeines zwischen Ansatz der 4. und der 5. Rippe . Rippen und Wirbelsaeule intakt .

Arbeitsplatz

Semantic Tagger

Ort	Beteiligte	Ereignis	Verletzung	Zeit

Sektionsergebnis In Schaedelhoehle links laengsverlaufende **Platzwunde** von bis auf die Schaedelschwarte . Geringgradige lackrote **Unterblutung** der Schaedelschwarte links und rechts hinten . **Bruch** des Schaedeldaches mit einer Bruchlinie , bogenfoermig nach oben verlaufend , im Bereich des Hinterhauptes mit einem Auslaeuer in Koerpermittellinie nach oben und vorn , etwa bis in Schaedelhoehle . Weiterer Verlauf der bogenfoermig verlaufenden Bruchlinie nach links und rechts seitlich zur Schaedelbasis , querverlaufend unmittelbar vor den Felsenbeinen . Scharnierartige Beweglichkeit der Bruchstuecken gegeneinander . **ZerreiSSung** der Hirnhaeute im Bereich der Schaedelbasis . **Durchtrennung** des Hirns im Bereich des Ueberganges vom Hirnstamm zum verlaengerten Rueckenmark . Abriss aller Brueckenvenen im Bereich der Hirnmantelkante . Ganz spaerliche **Blutung** unter die harte Hirnhaut . Zeichen des zentralen Todes und des erheblichen Blutverlustes : Blutaustritt aus beiden aeusseren Gehoergaengen bei Eroeffnung der Mittelohren und der Trommelfelle . Akute Erweiterung der Herzhoehlen . In den Herzhoehlen und grossen Gefaessen relativ wenig fluessiges Blut . Spaerliche Totenfleckbildung an der Koerperrueckseite . Blaesse und Hervortreten der Eigenfarbe der Organe . Blaesse und akute Ueberblaeung der Lungen . Schlaffe Erweiterung von Hamblase und Enddarm . Schraeg von links oben nach rechts unten verlaufende **Schuerfung** mit geringgradiger **Unterblutung** der Weichteile im Bereich der Stirnhaut und Schlaefenregion , links zur Jochbeinregion reichend . Ebensolehe Schuerfrichtung im Bereich des Nasenrueckens links und der Nasenspitze . Weiterhin ebensolehe **Schuerfung** im Bereich der Unterlippe und an der Rueckseite des linken Ohres . Keine Dehnungsrisse hinter dem linken Ohr feststellbar . Im Bereich der Halswirbelsaeule **Unterblutung** des vorderen Laengsbandes und Lockerung der Zwischenwirbelscheibe zwischen dem 4. und 5. Halswirbelkoerper . Seitenkantenabriss und **Unterblutung** des vorderen Laengsbandes im Bereich des 9. und 10. Brustwirbelkoerpers . Geringgradige **Blutung** in die Herznuskulatur im Bereich der linken Vorderwand . Maessige **Zerrung** im Bereich der Lungenwurzel beiderseits . **Zerrung** des Lippenbaendchens der Oberlippe und **Abspplitterung** an Anteilen des 1. und 2. Schneidezahnes im Unterkiefer rechts . Im Bereich des Mundbodens querverlaufende oberflaechliche **Hautabschuerfung** mit roetlicher Verfaerbung und **Weichteilschwellung** mit Betonung der rechten Seite von bis . Schraeg von links oben nach rechts unten verlaufende **Schuerfung** von x im Mittelbauchbereich rechts . Geringgradige **Weichteilblutung** im Bereich des rechten Armes . Oberflaechliche **Hautabschuerfung** an der Streckseite der rechten Hand . **Hautabschuerfung** und Ablederung vom Fettgewebe ueber der rechten Gesaessbacke . **Bruch** des Beckenringes rechts des rechten Kreuzbein-Darmbein-Gelenkes . Sprengung des linken Kreuzbein-Darmbein-Gelenkes . **Bruch** des linken oberen und unteren Schambeinastes neben der Schamfuge . Abriss des Gekroeses des Duennardmes beim Uebergang auf den Blinddarm auf - und im Bereich des sigmafoermigen Dickdarmes . KapselzerreiSSung im Bereich des oberen Poles der **Milz** bis zur Wurzel . Etwas Blut im Bereich der Bauchhoehle links . Deutliche lackrote **Weichteilunterblutung** um die Nieren , rechts staerker als links , mit **Zerrung** der Nierenstiele . An der Aussen- und Rueckseite des rechten Oberschenkels , ueber der rechten Kniescheibe , ueber dem Innen- und Aussenknoechel des rechten oberen Sprunggelenkes sowie ueber der Streckseite des rechten Fusses oberflaechliche , zumeist rillenfoermige **Hautabschuerfungen** . An der Innenseite des linken Oberschenkels , ueber der linken Kniescheibe , ueber der linken Schiebeinkante innen , an der Rueckseite des linken Unterschenkels und an der Aussenseite des linken Knoechels oberflaechliche **Hautabschuerfungen** mit teils geringgradigen **Unterblutungen** der Weichteile . Im unteren Lendenwirbelsaeulenbereich rechts **Weichteilunterblutung** . Abbruch der Spitze des Dornfortsatzes vom des 8. Brustwirbels . An der Innenseite beider Oberschenkel und unmittelbar unterhalb des Dammes mit Betonung der linken Seite sowie im Bereich des Dammes **Hautdefekte** und Weichteilblutungen . II . Chemisch-toxikologische Untersuchungen Die Bestimmung der Ethanolkonzentration erfolgte unter Beruecksichtigung des einschlaegigen Gutachtens des BGA und des Beschlusses des BGH vom 28. Juni und verlief im Oberschenkelvenenblut , Mageninhalt und Urin negativ . Demnach lag zum Zeitpunkt des Vorfalles keine Alkoholbeeinflussung vor . III . Todesursache Der Tod ist durch akutes zentrales Herz- und Kreislaufversagen bei offener Schaedel-Hirn-Verletzung mit Abtrennung des verlaengerten Halsmarkes vom Hirnstamm eingetreten . IV . Todesart Es handelt sich um eine nichtnatuerliche Todesart . Nach den vorliegenden Anknuepfungstatsachen wird von einem Unfallgeschehen ausgegangen . V . Vorgeschichte Am dd.mm.jj befuhr Herr X , gegen 19.50 Uhr , mit seinem PKW Opel Frontera , amtliches Kennzeichen ABC-CD xxx , die Bundesstrasse xxx X aus Richtung XX in Richtung YY . Vor ihm befand sich G 3/99 mit seinem Moped Simson S , um nach links in Richtung YZ abzubiegen . Hierbei kam es zur **Kollision** in der Form , dass der Opel Frontera von hinten auf das Moped auffuhr . Die Sitzbank verkeilte sich mit dem hinteren Anteil im Frontbereich des Moped Fahrs . Samen wurden am Moped auf dem Frontschutz links und im Bereich der Motorbank , unmittelbar vor dem Windschutzscheibe , in

Dokument: Übermittelt

Experiment Environment

- corpus: Medline abstracts
 - ca. 500 abstracts
 - ca. 90.000 tokens
- preprocessed
 - tokenizer
 - sentence splitter
 - tagger
 - maximum entropy POS tagger (PennTreebank Tagset++)
 - named entity recognition:
 - substances, enzymes, ...

Experiment Environment

<ABSTRACT>

...

<SENTENCE><DT>The</DT> <JJ>nuclear</JJ>
<NOUN>enzyme~~enzyme</NOUN> <NOUN>poly(ADP-
ribose)polymerase~~enzymeB</NOUN> <LPAR>(</LPAR>
<NOUN>PARP~~enzyme</NOUN> <RPAR>)</RPAR> <VBZ STEM =
"be">is</VBZ> <RB>also</RB> <VBN STEM = "cleave">cleaved</VBN>
<IN>upon</IN> <DT>the</DT> <NOUN>activation</NOUN> <IN>of</IN>
<NOUN>caspases~~enzyme</NOUN> <CMA>,</CMA> <CC>and</CC>
<NOUN>IETD~~substance</NOUN> <RB>similarly</RB> <VBZ STEM =
"prevent">prevents</VBZ> <NOUN>PARP~~enzyme</NOUN>
<NOUN>cleavage</NOUN> <SEPR>.</SEPR></SENTENCE>

...

</ABSTRACT>

- use external resources: Wordnet
 - lexical database
 - lexical relations, e.g. synonym, antonym
 - conceptual relations, e.g. hypernym, holonym, meronym

	Wordnet (1.7)
Noun	74488
Verb	12754
Adjective	18523
Adverb	3612
Total:	109377

Import of a Tagset

```
<WORDCLASS>
  <XDOC>*noun*</XDOC>
  <GERMAN>
    <TAG casus="cas" genus="gen"
  numerus="num">N</TAG>
  </GERMAN>
  <ENGLISH>
    <TAG
  numerus="SINGULAR">NOUN</TAG>
    <TAG
  numerus="PLURAL">NOUNS</TAG>
  </ENGLISH>
</WORDCLASS>
```

```
<WORDCLASS>
  <XDOC>*special*</XDOC>
  <GERMAN>
    <TAG>NE</TAG>
  </GERMAN>
  <ENGLISH>
    <TAG>SUBSTANCE</TAG>
    <TAG>ENZYME</TAG>
    <TAG>ENZYMEA</TAG>
    <TAG>ENZYMED</TAG>
  ...
  </ENGLISH>
</WORDCLASS>
...
```

Import of a Tagset

Generic Tags	German	English
noun	N	NOUN, NOUNS
verb	V	VB, VBN, VBZ, VBD, VBG, VBP
adverb	ADV	RB, RBS, RBR
adjective	ADJ	JJ, JJR, JJS, JJSS
special	NE	SUBSTANCE, ENZYME, ENZYMEDIA, ENZYMEB, ...

Results

package, package · ^{THIS MODEL} hypothesis, possibility, theory kind, sort, form, variety assistant, helper, help, supporter representation ideal
reader representation, mental representation, internal representation assistant, helper, help, supporter representation is distinct chiseled,
well-defined different (vs. same) separate (vs. joint) definite (vs. indefinite) clear (vs. unclear) from one proposed declare plan intend,
mean, think choose, take, select, pick out request for the related accompanying, attendant, concomitant, incidental, incidental to (predicate)
affinal similar (vs. dissimilar) relevant (vs. irrelevant) bacterium microorganism Klebsiella enteric bacteria, enterobacteria, enterics
pneumoniae, in which unmodified unadapted GlnK SUBSTANCE relieves NifL SUBSTANCE inhibition abstinence restraint, control
organic process, biological process action instead of stimulating arouse, elicit, enkindle, kindle, evoke, fire, raise, provoke affect effect,
effectuate, bring about, set up sensitize, sensitise challenge it. Streptococcus eubacteria, eubacterium, true bacteria salivarius is a
lactose- and galactose-positive bacterium microorganism that is phylogenetically phylogenetic, phyletic closely close (vs. distant)
close (vs. distant) close related think, cogitate, cerebrare tell, narrate, recount, spin, recite be interact associate, tie in, relate, link, colligate,
link up, connect to Streptococcus eubacteria, eubacterium, true bacteria thermophilus, a bacterium microorganism that metabolizes
lactose but not galactose. In this paper material, stuff essay press, public press article medium publisher, publishing house, publishing
we report

Results

Analysis of the crystals showed that they are formed by fully intact complex I in an **L-shaped** conformation. Activity assays and single particle analysis indicated that complex I maintains this structure in detergent solution and does not adopt a different conformation in the active state. Thus, we provide the first experimental evidence that complex I from *E. coli* has an **L-shape** in a lipid bilayer and confirm that this is also the case for the active enzyme in solution. This suggests strongly that bacterial complex I exists in an **L-shaped** conformation in vivo. Our results also indicate that native lipids play an important role in the **activation**, stabilization and, as a consequence, crystallization of purified complex I from *E. coli*. The product of oxygenation of arachidonic acid by the **prostaglandin H synthases (PGHS)**, prostaglandin (PG) **H2**, undergoes rearrangement to the highly reactive α -ketoaldehydes, levuglandin (LG) **E2** and **LGD2**. We previously have demonstrated that **LGE2** reacts with the ϵ -amine of lysine to form both the levuglandinyl-lysine Schiff base and the pyrrole-derived levuglandinyl-lysine lactam adducts. We also have reported that these levuglandinyl-lysine adducts are formed on purified **PGHSs** following the oxygenation of arachidonic acid. We now present evidence that the levuglandinyl-lysine lactam adduct is formed in human platelets upon **activation** with exogenous arachidonic acid or **thrombin**. After proteolytic digestion of the **platelet** proteins, and isolation of the adducted amino acid residues, this adduct was identified by liquid chromatography-tandem mass spectrometry. We also demonstrate that formation of these adducts is inhibited by indomethacin, a **PGHS inhibitor**, and is enhanced by an **inhibitor** of **thromboxane synthase**. These data establish that levuglandinyl-lysine adducts are formed via a **PGHS-dependent pathway** in whole cells, even in the presence of an **enzyme** that metabolizes **PGH2**. They also demonstrate that a physiological stimulus is sufficient to lead to the lipid modification of proteins through the levuglandin **pathway** in human platelets. Yeast **cytosine deaminase** is an attractive candidate for anticancer gene therapy because of its catalysis of the deamination of the prodrug 5-fluorocytosine to form 5-fluorouracil. We report here the crystal structure of the **enzyme** in complex with the **inhibitor** 2-hydroxypropylidene at 1.6

Results: Matches in Wordnet

	total	WordNet
Nouns	32211	23955
Verb	11160	5716
Adjective	9710	6618
Adverb	2528	1581
total:	55609	37870

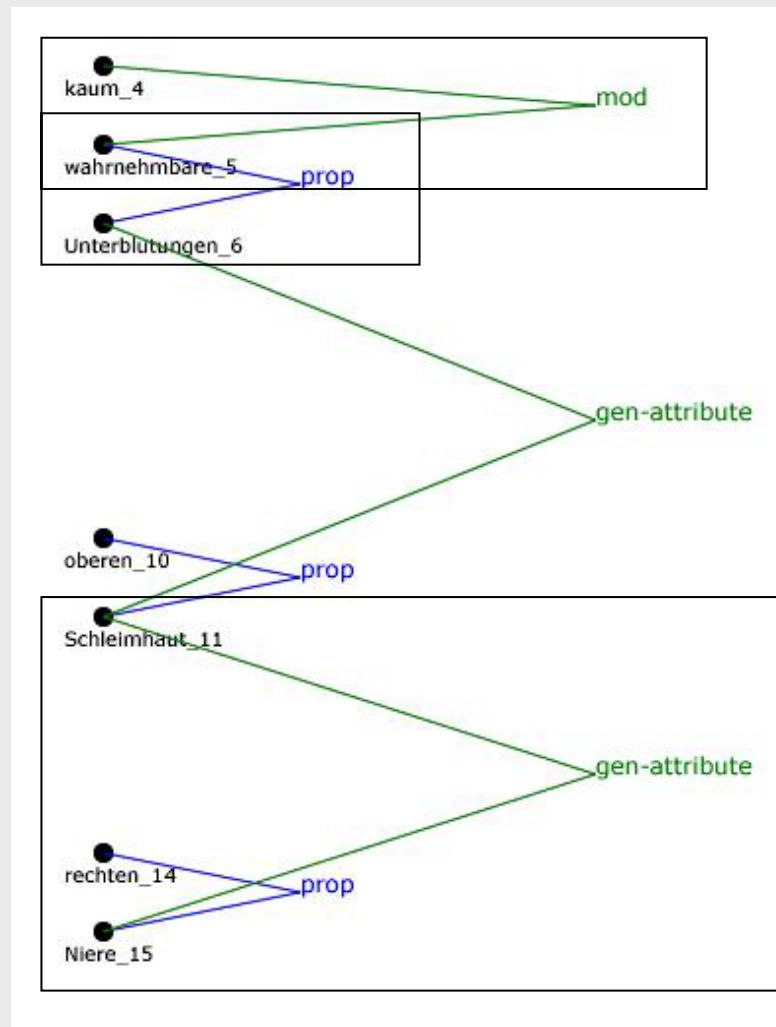
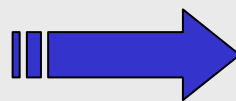
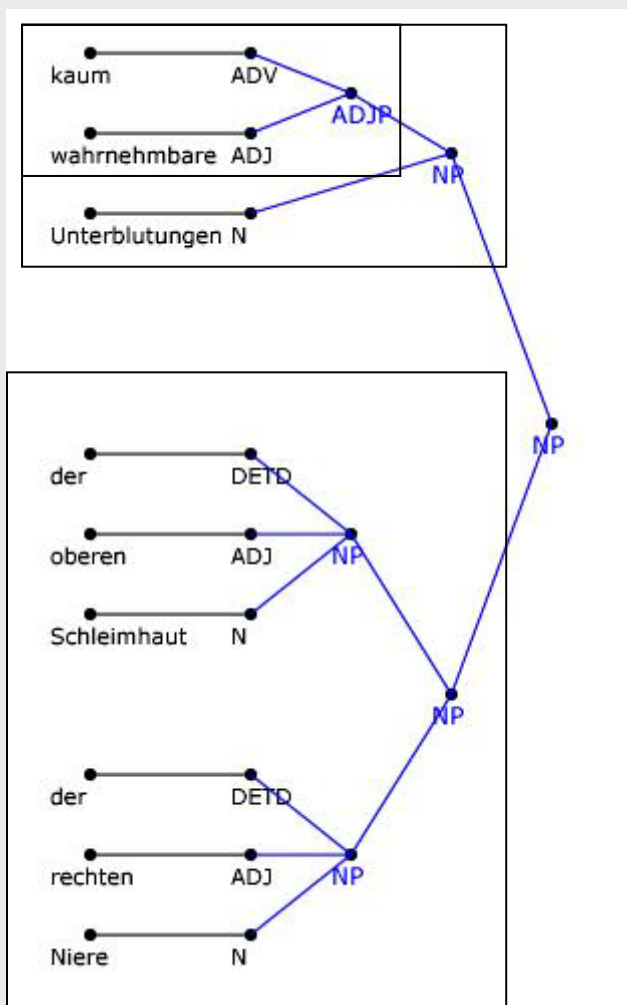
- not tested: 41172

- processing Medline abstracts
 - coverage > 50%
- problems
 - errors by preprocessing
 - POS tagger
 - tokenizer
 - input data: not XML conform
 - elements like ... „*sandwich*“... or ... >1,5 mol ...

Next Steps: SIsS

- evaluation of SIsS
- SIsS: **S**emantic **I**nterpretation of **s**yntactic **S**tructures
- approach:
 - input data: parsed syntactic structures
 - search for predefined structures (grammar rule)
 - semantic analysis of this structure
- resource:
 - mapping of grammar rules to an interpretation
 - core of a rule
 - attributes for the core

SIS: from Syntactic to Semantic



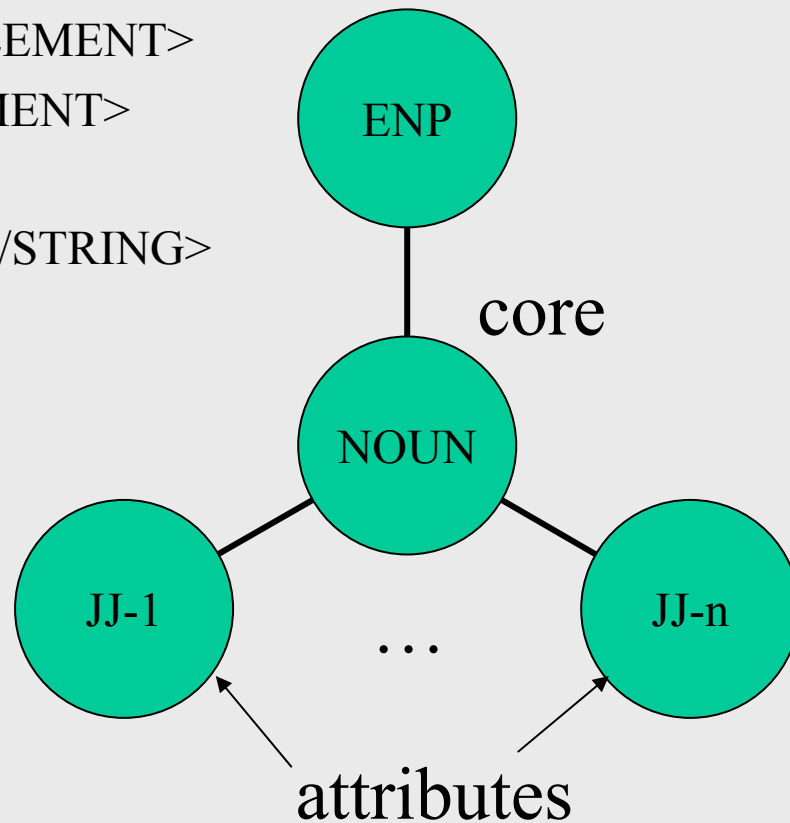
in English: hardly detected bleedings of upper mucous membrane of right kidney

SIsS for English

- example:

```
<RULE RULE-CAT="ENP"  RULE="NP2">
  <ELEMENT>DT</ELEMENT>
  <ELEMENT OP="*">JJ</ELEMENT>
  <ELEMENT>NOUN</ELEMENT>
  <EXAMPLE>
    <STRING>the pretty girl</STRING>
  </EXAMPLE>
</RULE>
```

NP2 → DT JJ* NOUN

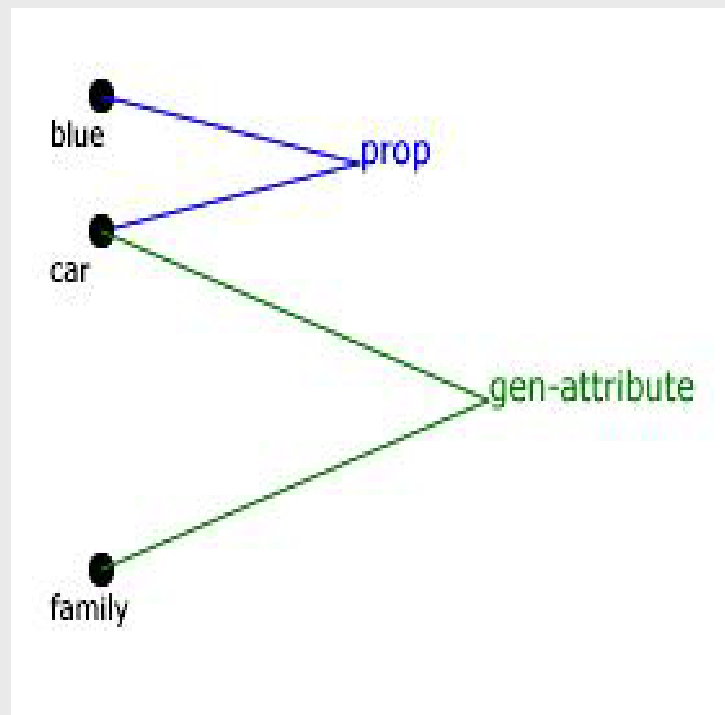
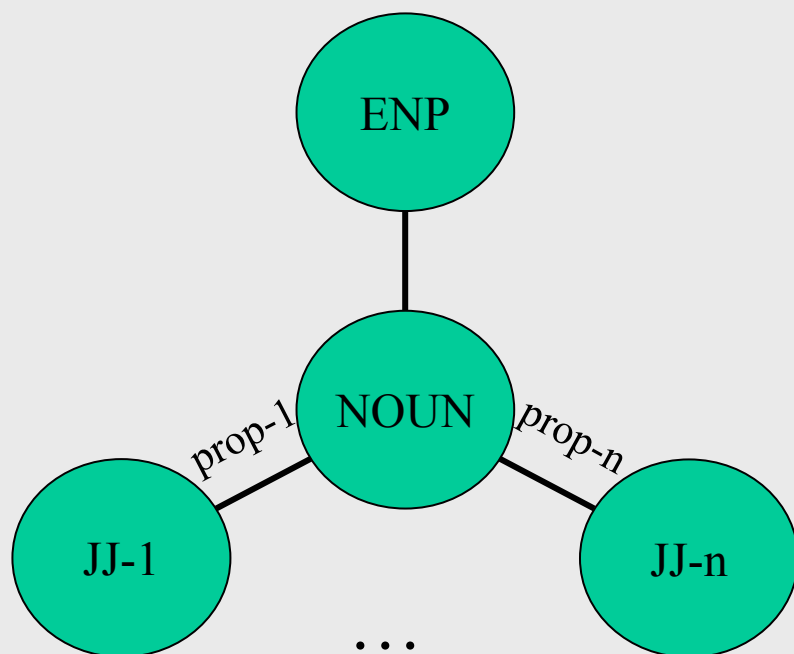


SIsS: Resources

- wordclasses:
 - concepts: NOUN, NOUNS, JJ, JJS
 - relations:
 - JJ, JJS: property
 - RB,: modifier
- grammar rules:
 - NP2: NOUN (core)
 - or
 - NP34: NP (core), relation: ‚genitive-attribute‘
NP34 → NP PP (of)

SIS: Results

results: prop-1(JJ-1, NOUN), ..., prop-n(JJ-n, NOUN)



example: the blue car of the family ...

- SIsS:
 - use of WordNet's holonym and meronym relation for resolving of 'genitive-attribute' relation
- Semantic Parser:
 - use of WordNet's verb frames for 'shallow' case frame analysis
 - on the base of semantic fields

Questions ?