Towards Knowledge-Based Assistance for Scholarly Editing

Jana Kittelmann
MLU Halle-Wittenberg

Christoph Wernhard
TU Dresden

AITP 2016

Obergurgl, 6 April 2016

Extended version of the talk slides, 19 April 2016

1. Scholarly Editing
2. Relevant Knowledge Sources
3. KBSET – An Experimental Platform
4. Coupling Fuzzy and Symbolic Knowledge
5. Access Predicates
6. Conclusion

1. **Scholarly Editing**
2. Relevant Knowledge Sources
3. KBSET – An Experimental Platform
4. Coupling Fuzzy and Symbolic Knowledge
5. Access Predicates
6. Conclusion

Scholarly Editing as Scientific Discipline

- Some other/related names/concepts:
  - Editionswissenschaft, Editionsphilologie, Editorik
  - Critique génétique
  - Textual criticism
- Emerged in the 1850s from reconstruction of ancient and medieval texts
- Outcome: **critical edition**
- Concerns
  - tracing and presenting text genesis
  - identifying a "definitive" version
  - presentation
  - bridging temporal and cultural distance to reader
  - "objective editions are not possible"

Summary Editions (Regestausgaben) of Correspondences

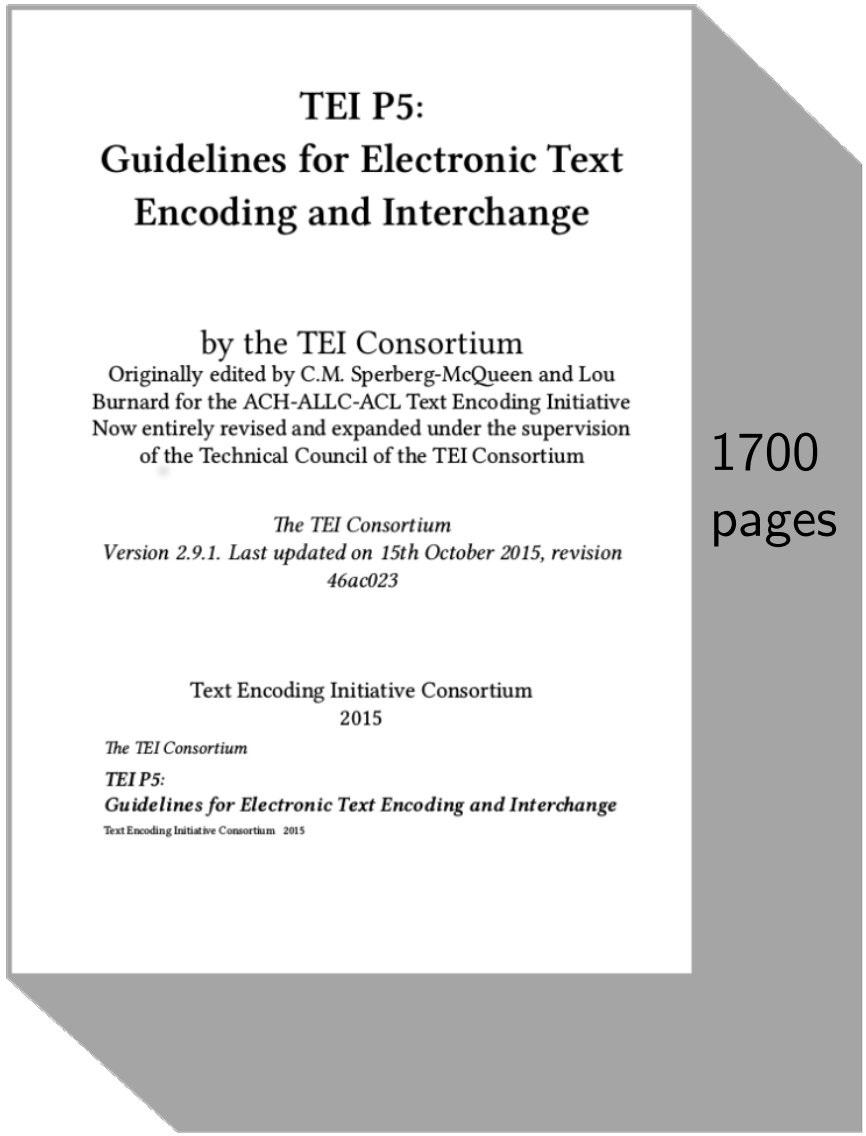
- Cases with **too much material to transcribe and present in full**

  Example: 20.000 letters to Goethe – successively published since the 1980s

- "Flat" forms of making accessible
  - involved persons
  - locations
  - dates
  - mentioned works
  - historic events
  - indexes

Separation of Descriptive and Procedural Markup: TEI

- Specification of XML elements and attributes for descriptive markup



TEI: Example

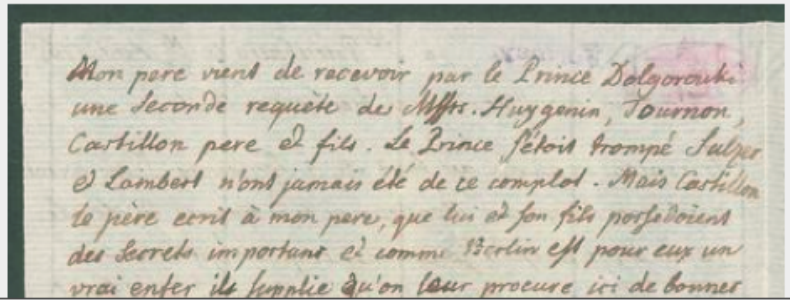
emacs

File Edit Options Buffers Tools TEI UniChar XML Help

```
      >Bouc</persName> est ici et loge <lb/>à <placeName>Wasiliostrov</placeName>
  avec un Officier hollandois ; il se porte <lb/>bien à ce qu'on m'a dit ; je n'ai
  pas encore pû le <lb/>voir, et m'informer de son état et de ce qu'il fait.
  </p><pb n="2" facs="00000067.jpg"/><p><rs ref="#p0037">Mon père</rs> vient de
  recevoir par le <persName ref="#p2508">Prince Dolgorouki</persName>
  <lb/>une Seconde requête des Mssrs. <persName ref="#p2509">Huggenin</persName>,
      <persName ref="#p2548">Tournon</persName>, <lb/><persName ref="#p2511"
      >Castillon</persName> pere et fils. Le Prince s'étoit trompé <persName
      ref="#p2510">Sulzer</persName>
  <lb/>et <persName ref="#p2360">Lambert</persName> n'ont jamais été de ce
  complot. Mais <persName ref="#p2511">Castillon</persName>
  <lb/>le père ecrit à <rs ref="#p0037">mon pere</rs>, que lui et son fils
  possedoient <lb/>des Secrets importans et comme <placeName ref="#l1103"
      >Berlin</placeName> est pour eux un <lb/><hi rend="underlined">vrai
      enfer</hi> il supplie qu'on leur procure ici de bonnes <lb/>places etc. <rs
      ref="#p0037">mon pere</rs> a envoié leur requete au <persName ref="#2340">com
-&U:---  Brief24vonJAEuleranJHSFormey.xml   52% (236,37)  Git@master  (nXML Invalid Fil
```

Vollbild zweite Spalte schließen

PDF-Download S. 1 2 3 · 5 vorheriger Brief nächster Brief



Mon père vient de recevoir par le Prince Dolgorouki une Seconde requête des Mssrs. Huggenin, Tournon, Castillon pere et fils. Le Prince s'étoit trompé Sulzer et Lambert n'ont jamais été de ce complot. Mais Castillon le père ecrit à mon pere, que lui et son fils possedoient des Secrets importans et comme Berlin est pour eux un vrai enfer il supplie qu'on leur procure ici de bonnes places etc. mon pere a envoié leur requete au comte d'Orlov mais je doute qu'on y fasse réflexion. Le Prof. Lowitz

Faksimile Dipl. Umschrift Lesefassung Metadaten Entitäten XML

Faksimile Dipl. Umschrift Lesefassung Metadaten Entitäten XML

TEI: Remarks

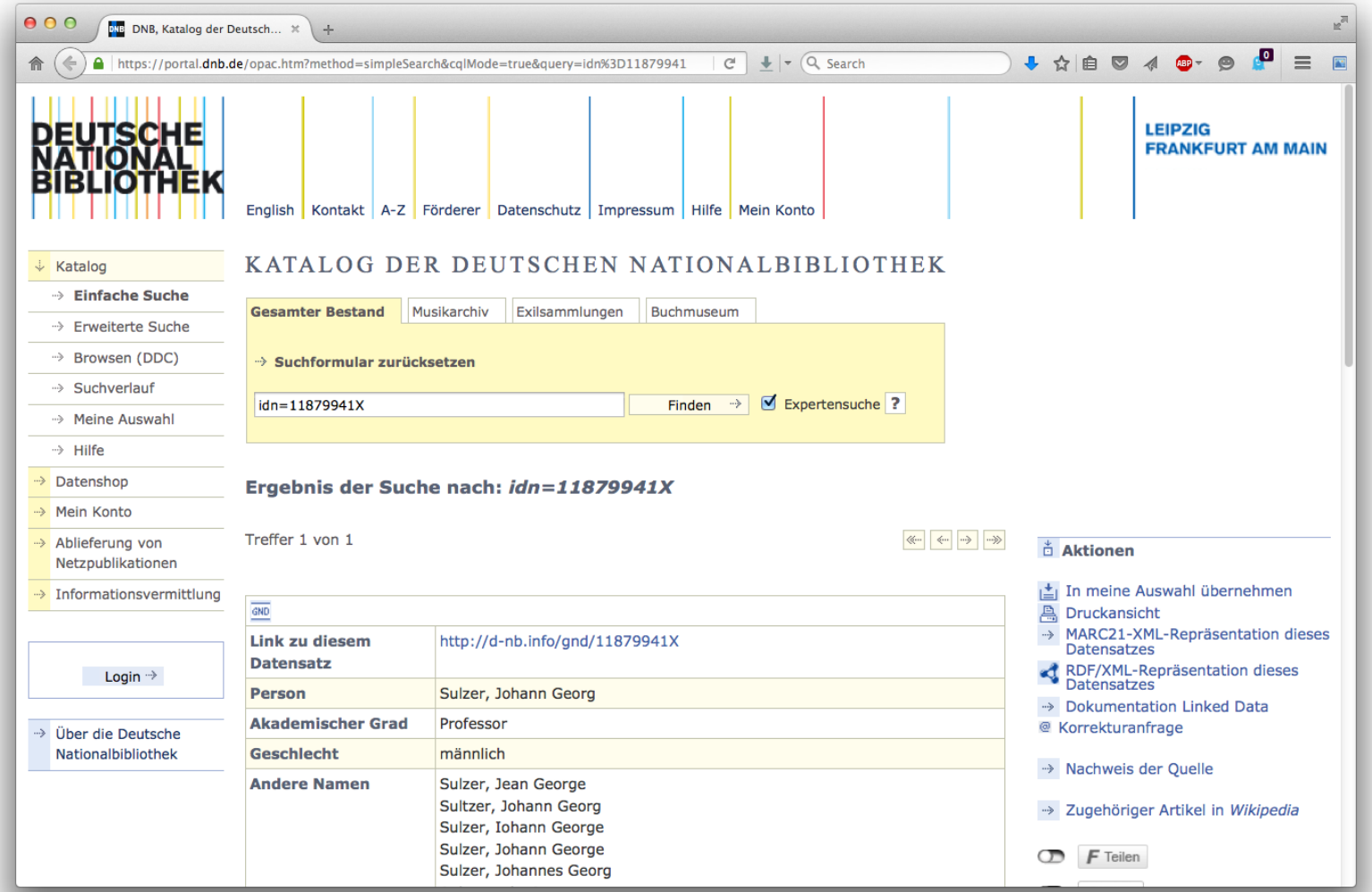
- TEI P5 2.9.2 (2015) `<correspDesc>`
- TEI P5 (2007) Entity descriptions: `<person>`, `<place>`, `<date>`
- Stand-off markup with W3C XInclude

1. Scholarly Editing
2. **Relevant Knowledge Sources**
3. KBSET – An Experimental Platform
4. Coupling Fuzzy and Symbolic Knowledge
5. Access Predicates
6. Conclusion

Wikipedia, Wikidata

Gemeinsame Normdatei [“Common Authority File”] (GND)

- Persons, organizations, works, . . .
- 3 M persons, 120 M facts
- Ontology with 60 classes
- Free (CC0)
- 10 GB RDF



GND Example

Weitere Angaben	Professor der Mathematik; 1775 Direktor der phil. Klasse der Akademie der Wisse
Beziehungen zu Personen	Sulzer-Keusenhoff, V Keusenhof, Johann A Graff, Elisabetha Sop
Systematik	4.7p Personen zu Ph
Typ	Person (piz)
Autor von	13 Publikationen 1. *Kurzer Begriff Sulzer, Johan* 2. *Versuch von d Sulzer, Johan der Ausg. Zür von Olaf Breid* 3. ...
Beteiligt an	5 Publikationen 1. *"Mein lieber H Winterthur : S* 2. *Allgemeine Th Ressource] Berlin : Direct* 3. ...
Thema in	1 Publikation 1. *Johann Georg Berlin : Akad.*

GND

Link zu diesem Datensatz	http://d-nb.info/gnd/11879941X
Person	Sulzer, Johann Georg
Akademischer Grad	Professor
Geschlecht	männlich
Andere Namen	Sulzer, Jean George Sultzer, Johann Georg Sulzer, Iohann George Sulzer, Johann George Sulzer, Johannes Georg Sulzer, Ioh. George Sulzer, Johann G. Sulzer, ... Sulzer, Jean G.
Quelle	M; B 1996; ADB 37 (1894), S. 144-147 M; B 1996
Zeit	Lebensdaten: 1720-1779
Land	Schweiz (XA-CH); Deutschland (XA-DE)
Geografischer Bezug	Geburtsort: Winterthur Sterbeort: Berlin Wirkungsort: Magdeburg Wirkungsort: Berlin
Beruf(e)	Philosoph Pädagoge Theologe Mathematiker Hochschullehrer

GeoNames

- 2.8 M locations, 10 M names
- Free (CC-BY)
- Table format

YAGO, DBPedia

- **Combined fact bases** from Wikipedia, GeoNames, ...
- Developed in computer science
- 5–10 M Objects, 100-3000 M facts
- 700–350.000 classes, based on Wikipedia and WordNet
- Mulit-lingual
- Free licenses
- RDF

1. Scholarly Editing
2. Relevant Knowledge Sources
3. **KBSET – An Experimental Platform**
4. Coupling Fuzzy and Symbolic Knowledge
5. Access Predicates
6. Conclusion

Addressed Issues in Scholarly Editing

- Incorporation of **automated techniques**, e.g.
  - named entity identification
  - statistics-based methods for analysis
- Providing explicit relationship to
  - **external knowledge bases**
  - **formal semantics**
- **High-quality presentations**
  - without expensive transformations and stylesheets
- **Loose coupling of object text and markup**
  - markup by different authors
  - automatically generated markup

Some AI Aspects Reflected in Scholarly Editing

AI	SE
• General background knowledge	• GND, GeoNames
• Position of the agent in the environment	• Position in the text
• Temporal order	• Order of word occurrences
• Incompletely sensed/understood environment	• Incompletely understood text
• Coming to decisions about actions to take	• Coming to decisions about denotations of phrases, about annotations to insert

The KBSET System

- "**K**nowledge-**B**ased Support for **S**cholarly **E**diting and **T**ext Processing"
- **Free software**: GNU Public License
- With comprehensive **example** (draft)

  Max Stirner: *Geschichte der Reaction*, Vol. 1, 1852

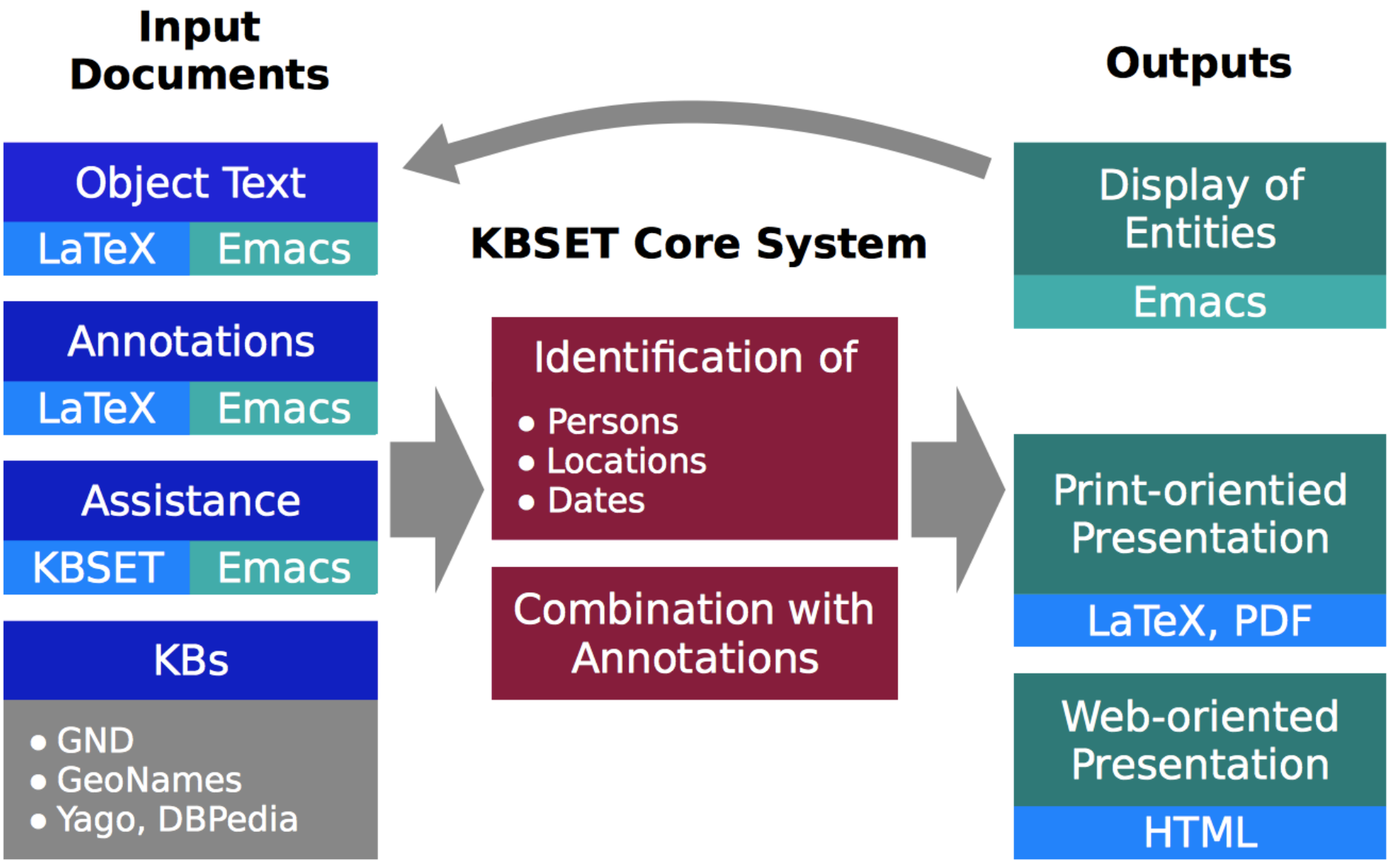
Guiding Principles

- All **phases** of editing should be supported
  1) **Creating the extended object text**
  2) Generating **intermediate representations** for examination by humans or machines
  3) Generating **final presentations**

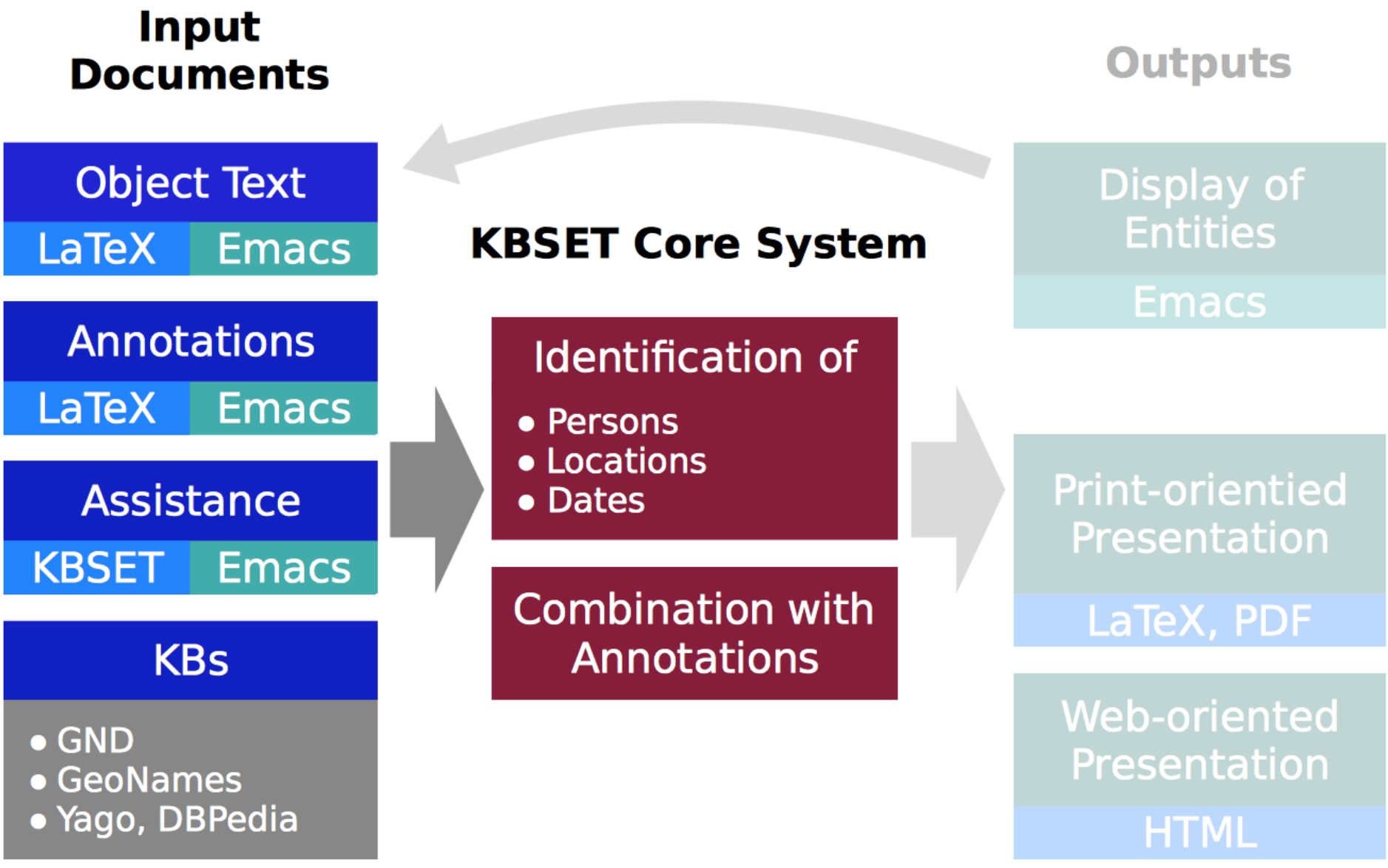
- **High quality** is required for all phases, e.g.
  - good tools for text creation
  - precisely identified persons
  - professional layout

- Consequences:
  - incorporation of **special techniques** and **special systems**
  - **automated techniques, adjustable by humans**

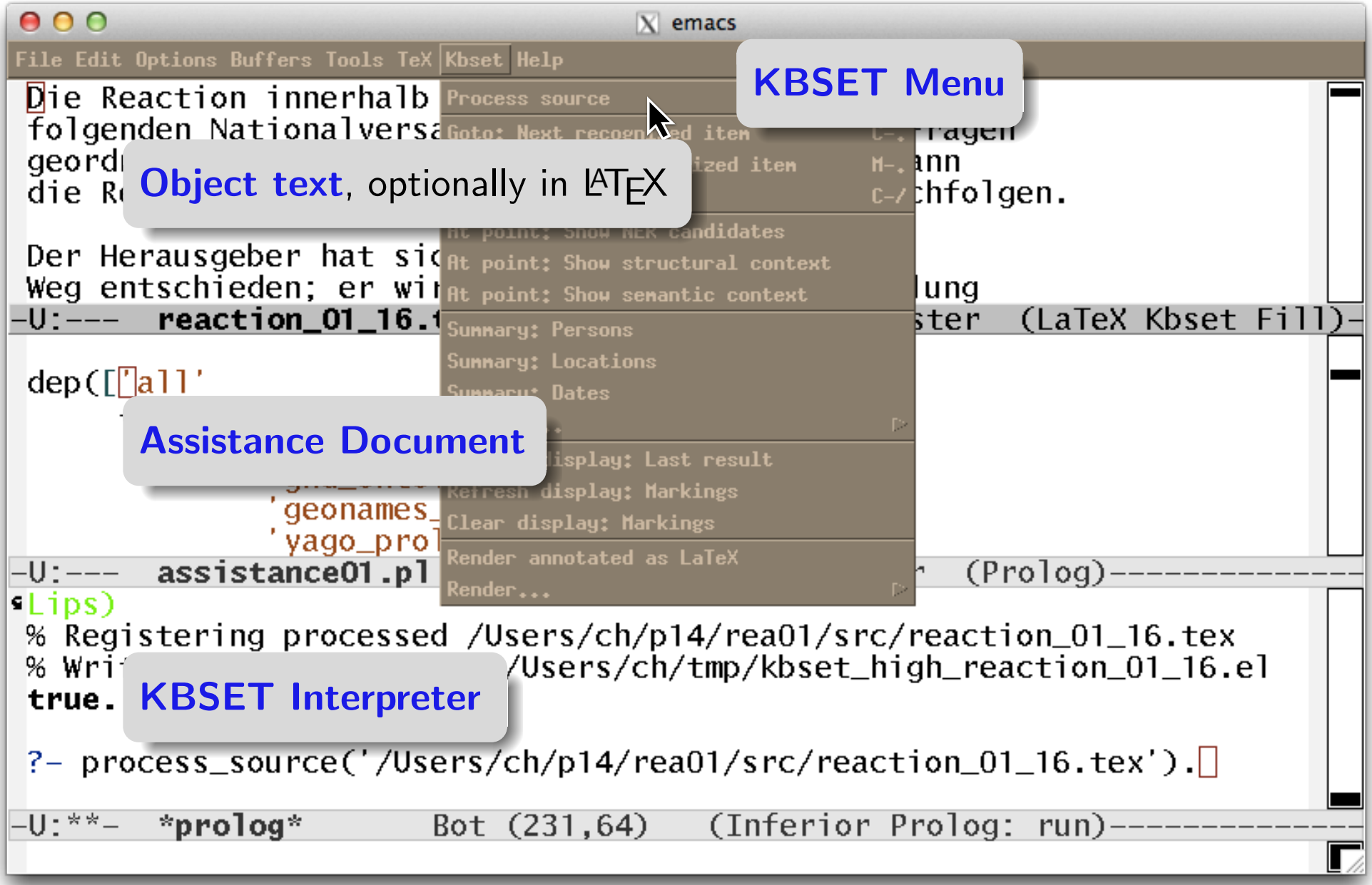
Overview


Input Documents
Object Text
LaTeX
Emacs
Annotations
LaTeX
Emacs
Assistance
KBSET
Emacs
KBs
• GND
• GeoNames
• Yago, DBPedia
KBSET Core System
Identification of
• Persons
• Locations
• Dates
Combination with Annotations
Outputs
Display of Entities
Emacs
Print-orientied Presentation
LaTeX, PDF
Web-oriented Presentation
HTML

Processing of Inputs



Embedding into Emacs



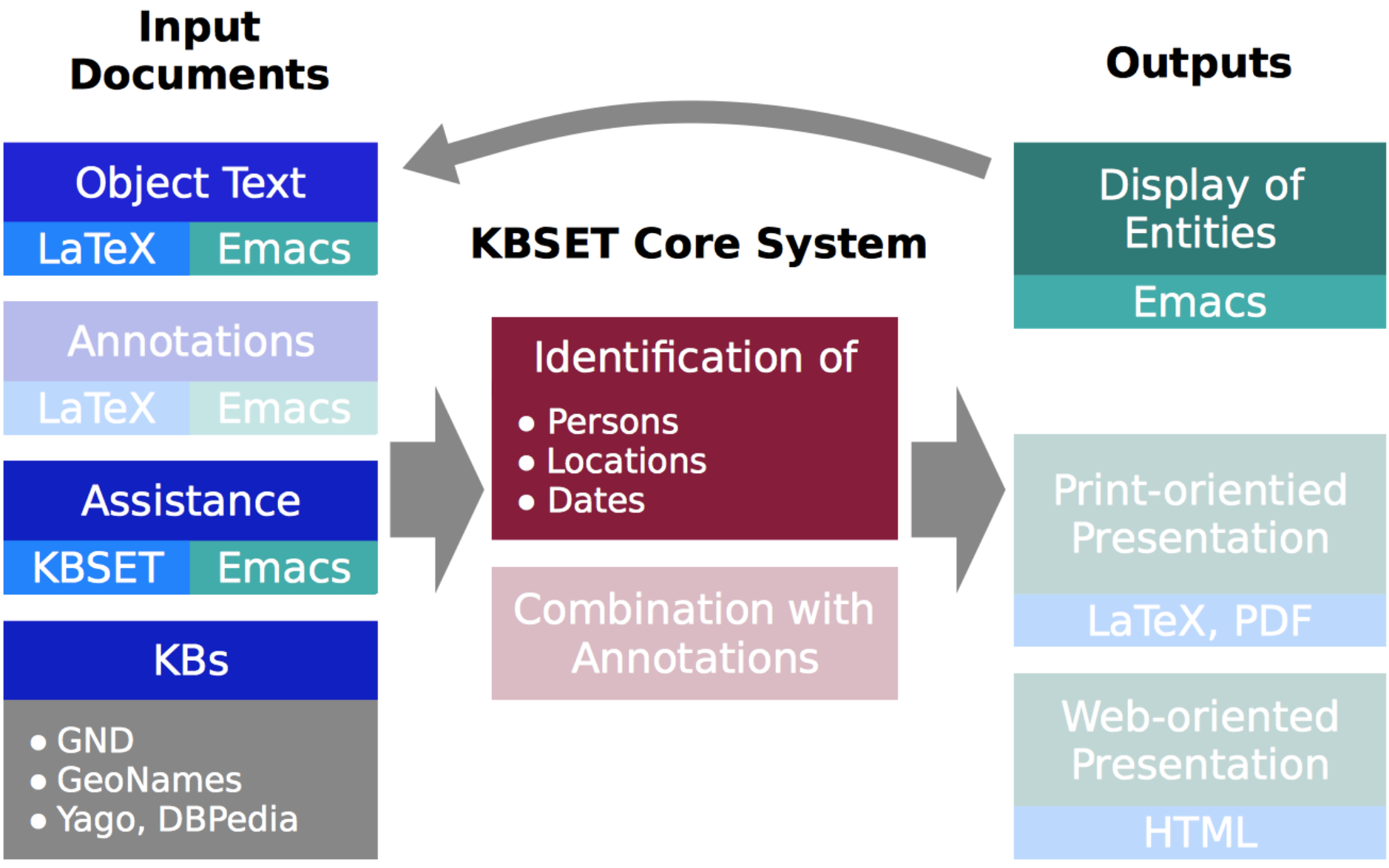
System Perspective on Knowledge Bases

- KBSET is implemented in **SWI-Prolog**
- . . . with theorem provers in mind, but currently making substantial use of
  - set abstraction (`findall`, `setof`)
  - sorting by term order
  - indexing on first argument
- **Preprocessing** for efficient access
  - **extracting** relevant data
    - GND: persons born before 1850 – 420 k instead of 3 M
  - indexed **access predicates**

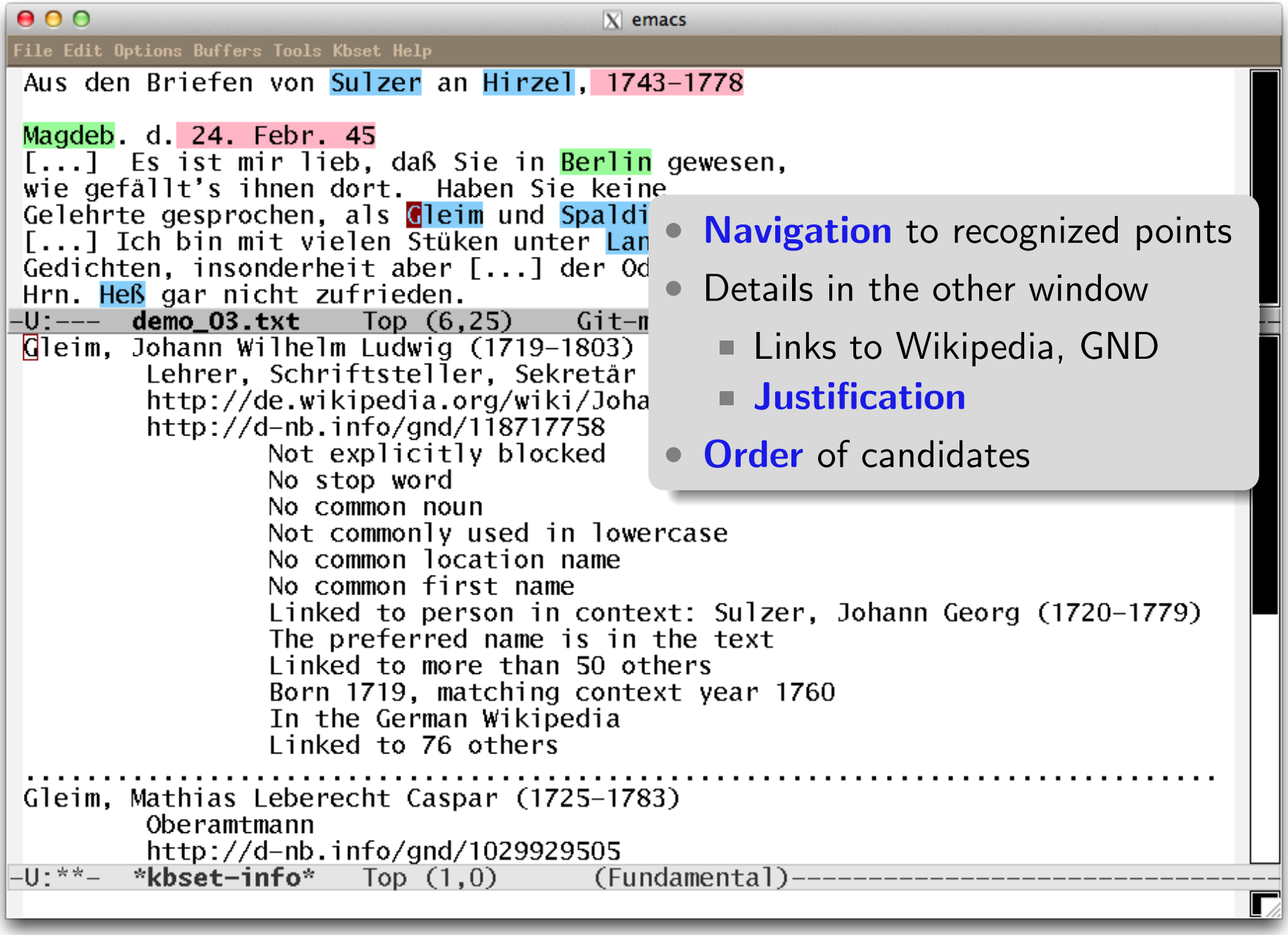
System Perspective on Text Representation

- **Sequence of units**: word | space | punctuation | command
  - allow to associate information, e.g. about identified entities
  - mapping to/from sequence of characters

Entity Identification

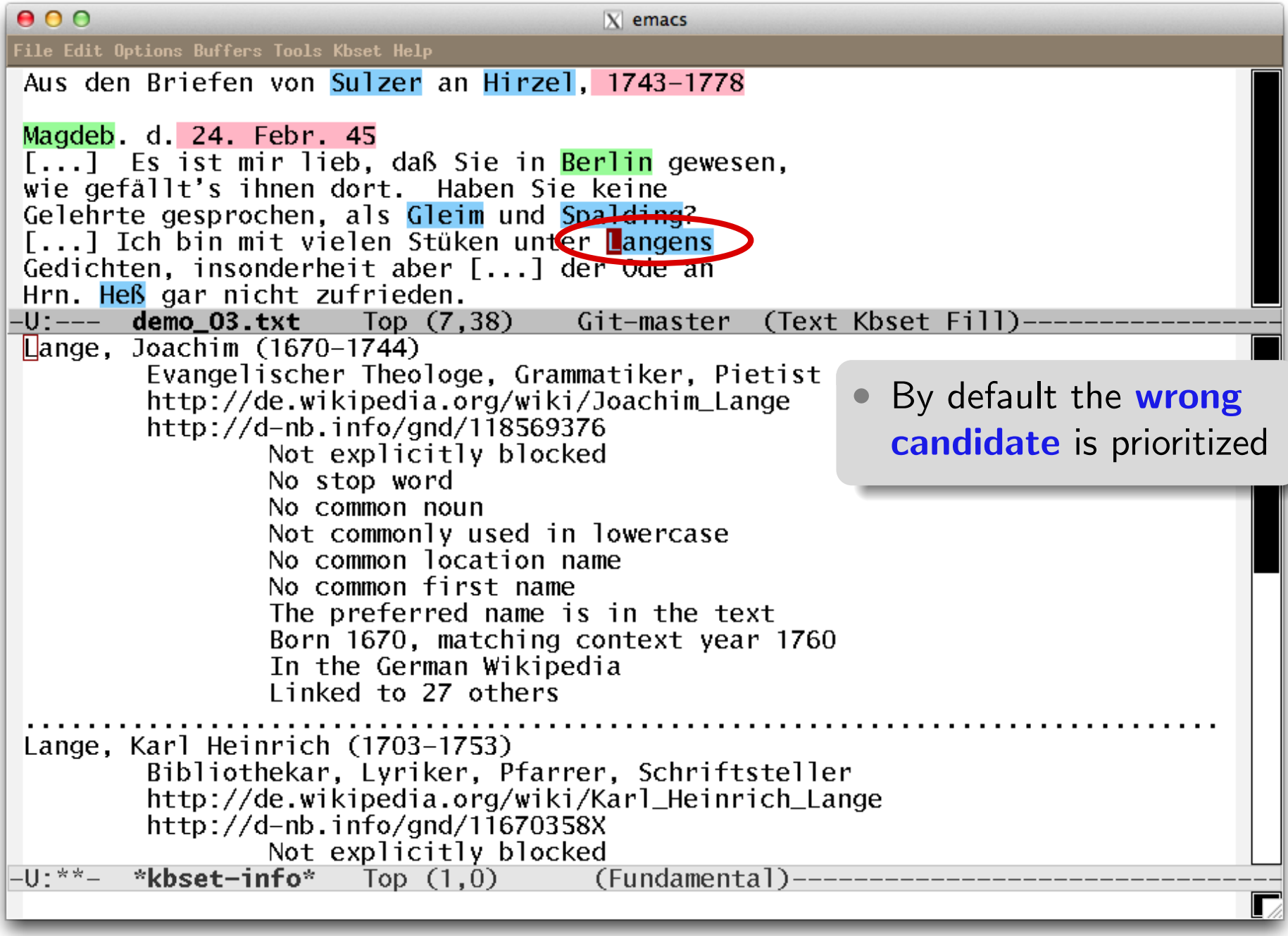


Identification of Persons



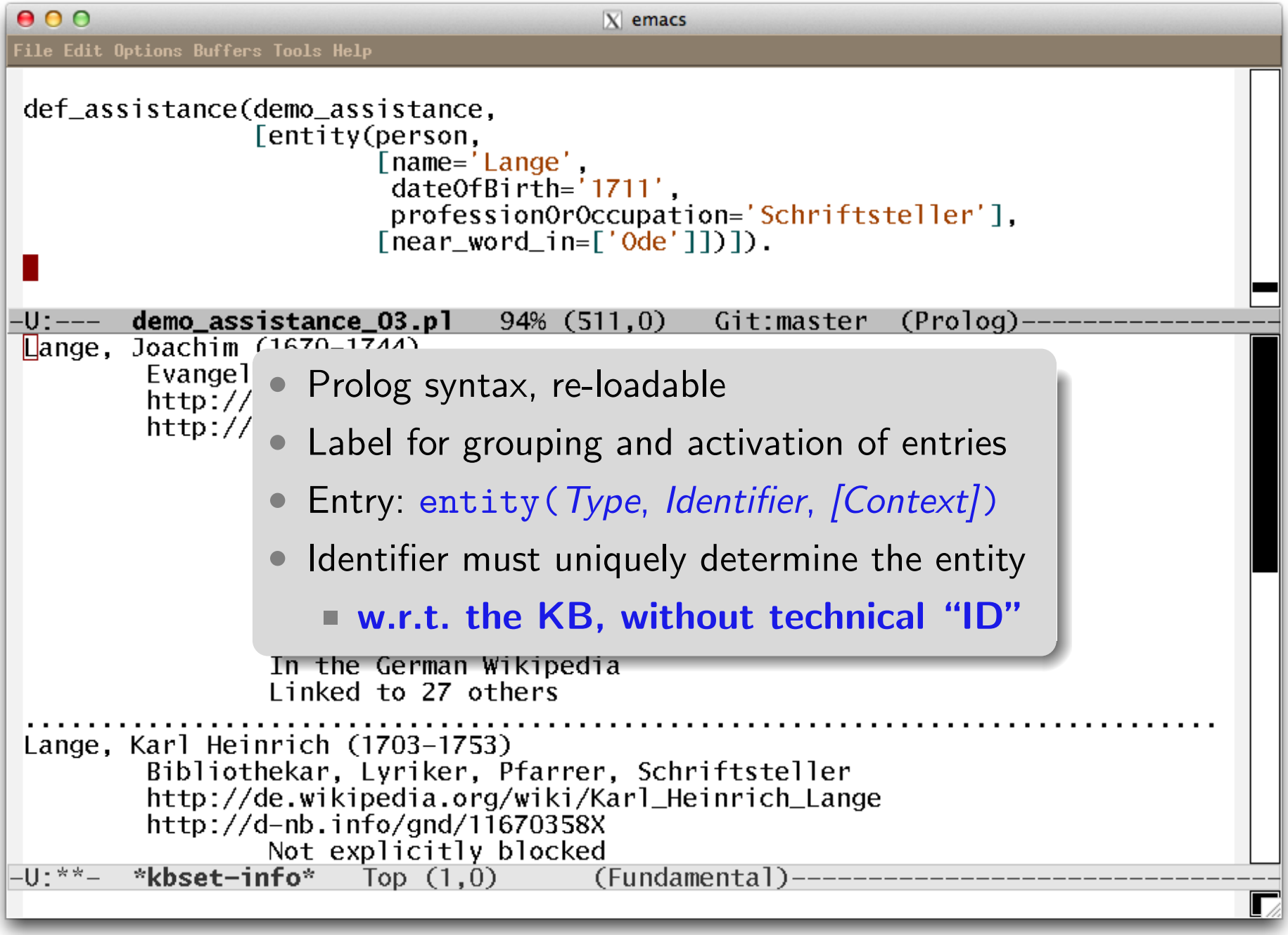
- **Navigation** to recognized points
- Details in the other window
  - Links to Wikipedia, GND
  - **Justification**
- **Order** of candidates

"Assistance" is Required Here



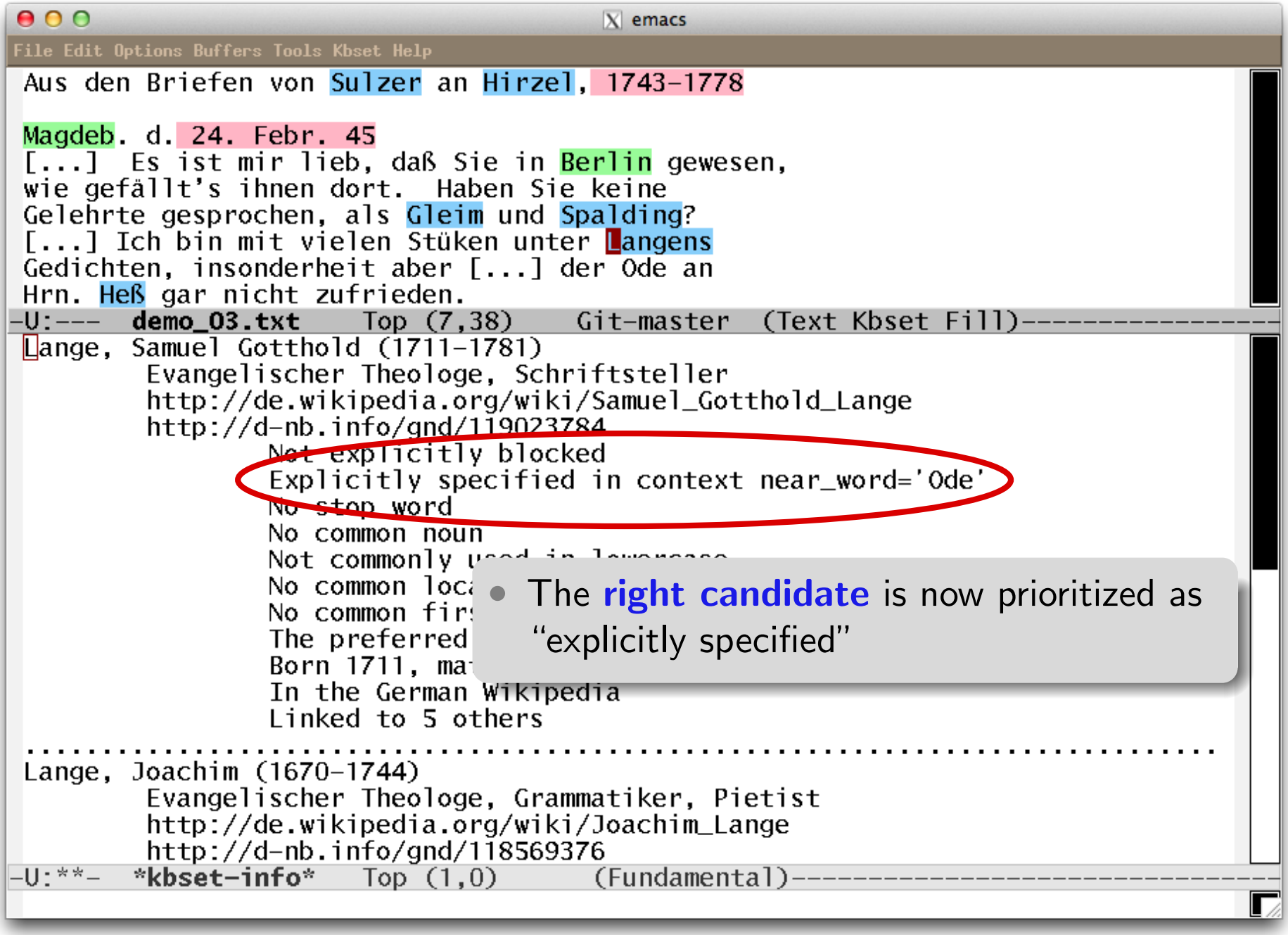
- By default the **wrong candidate** is prioritized

Entry in the Assistance Document



- Prolog syntax, re-loadable
- Label for grouping and activation of entries
- Entry: `entity(`*Type*, *Identifier*, *[Context]*`)`
- Identifier must uniquely determine the entity
  - **w.r.t. the KB, without technical “ID”**

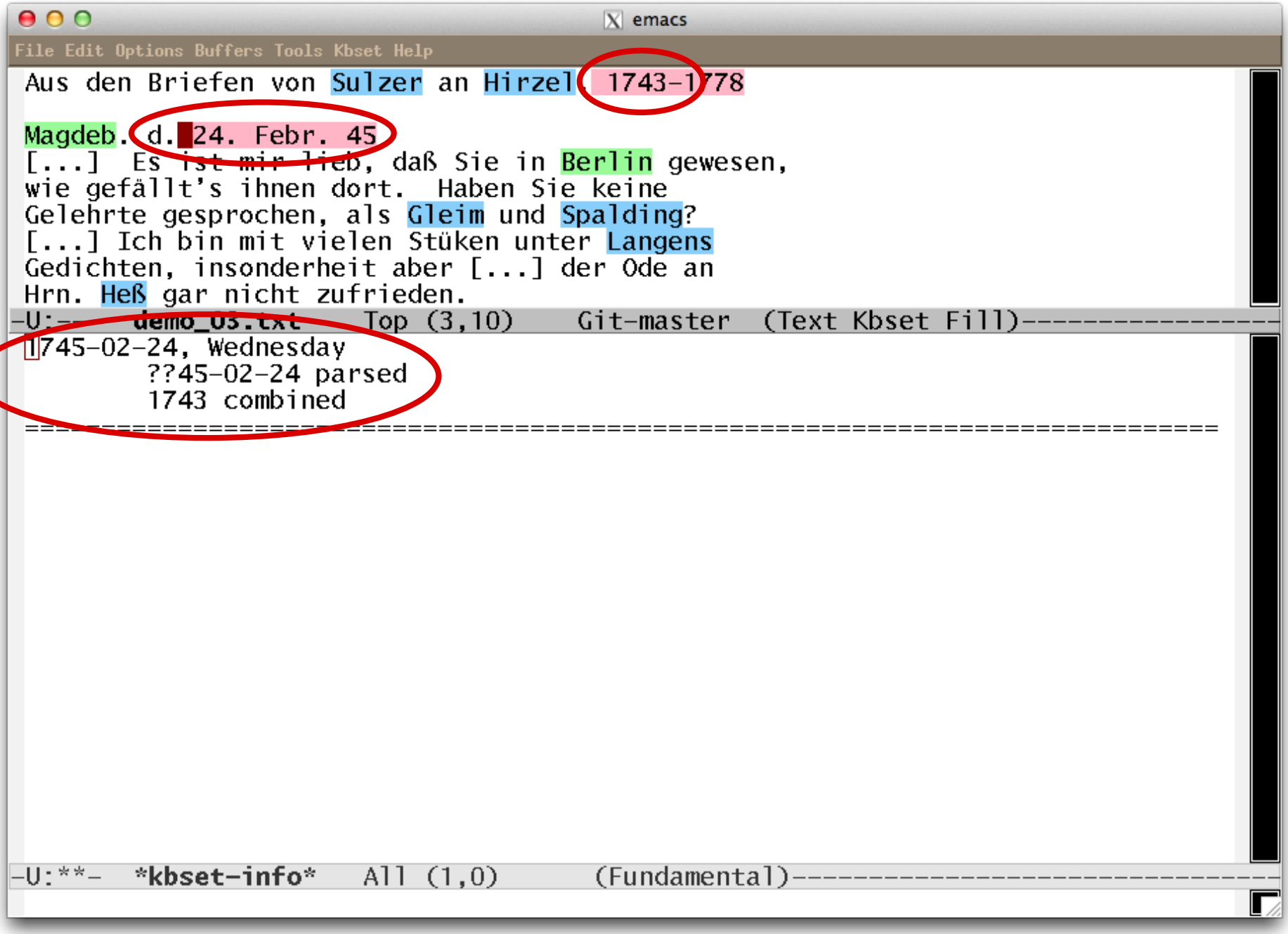
Correction after Adaption by “Assistance”



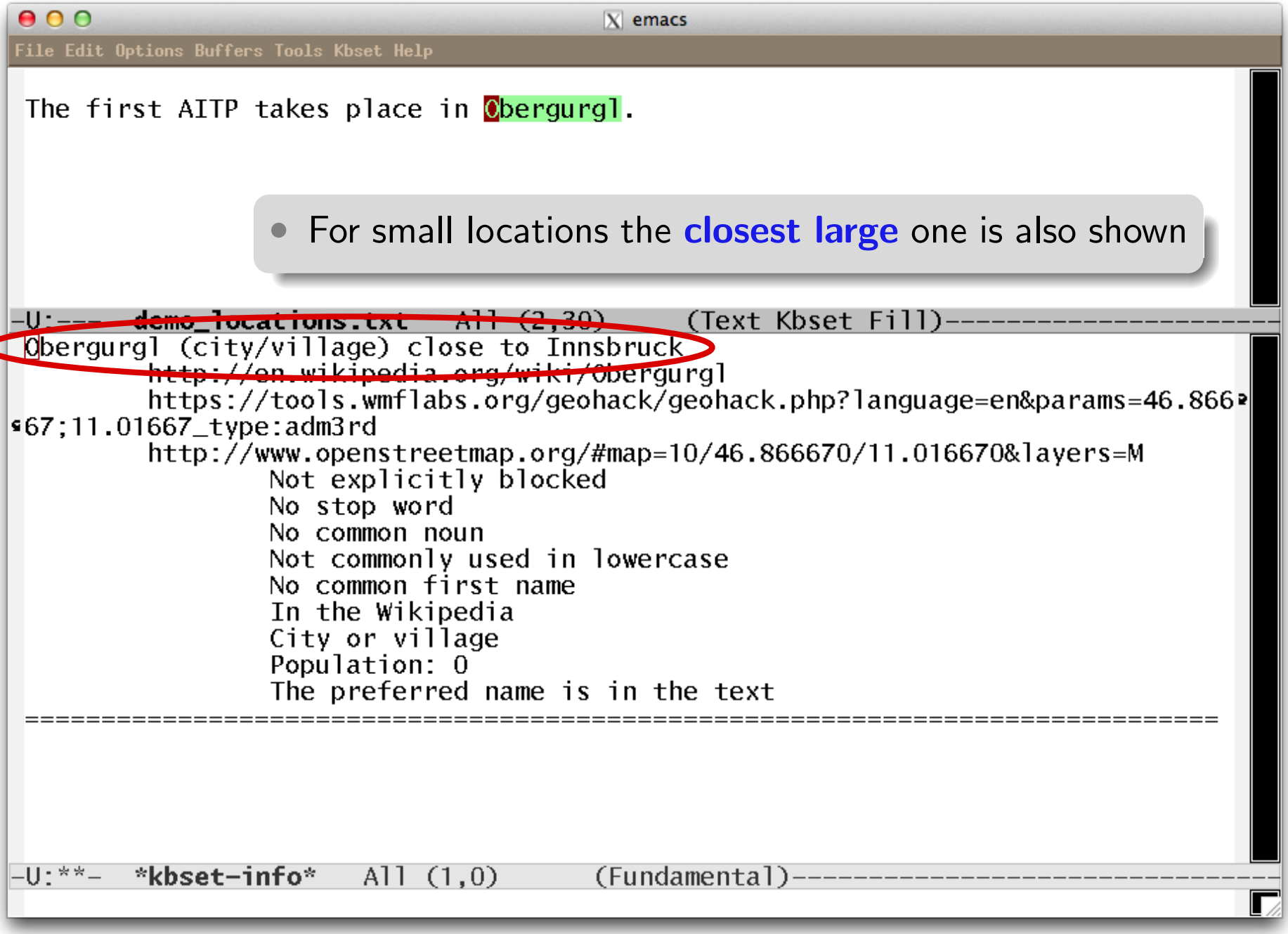
Further Possibilities in Assistance Documents

- Supplementing
  - **attribute values**
  - **entities**
- **Excluding words as entity designators**

Dates: Parsing and Defaulting



Detailed Information on Locations



Associated with Occurrences of Words

- In contrast to n-grams (sequences) of words
- **Local context** is considered
  - preceding and succeeding words
  - already identified entities

Comparison with a Popular Entity Recognizer

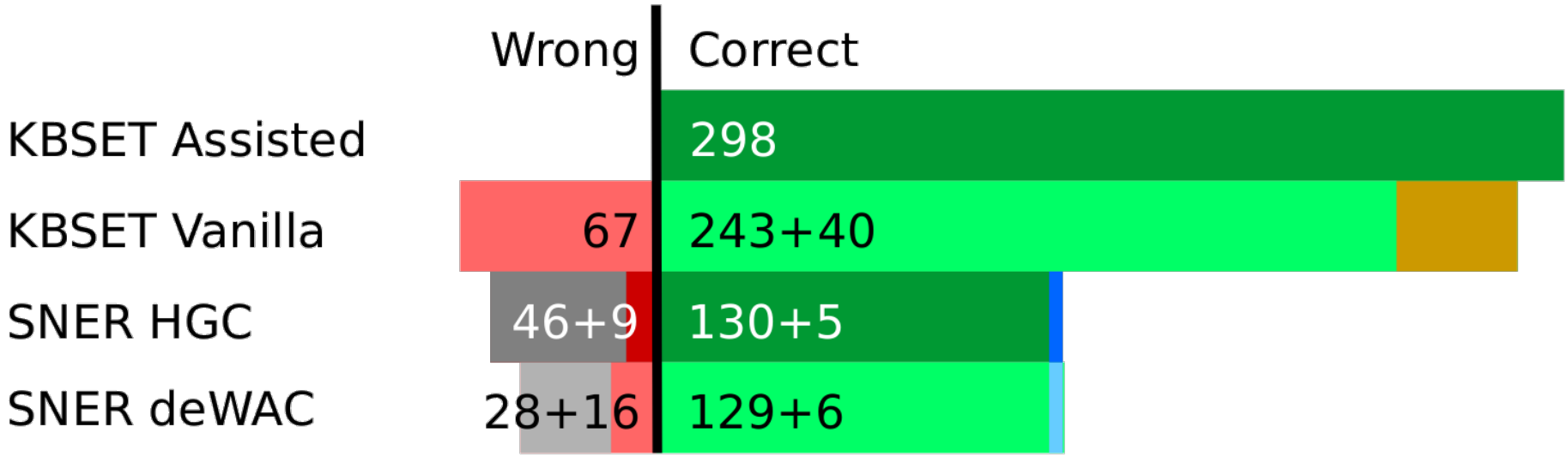
- **Stanford Named Entity Recognizer**
  - **statistics-based machine learning** [Finkel et al., 2005]
  - free, since 2006, here version 3.3.1 (Jan 2014)
  - **no identification, just recognizing the entity type!**

```
... in/O Berlin/I-LOC gewesen/O,/O wie/O
gefällt/O's/O ihnen/O dort/O./O Haben/O Sie/O keine/O
Gelehrte/O gesprochen/O,/O als/O Gleim/I-PER und/O
Spalding/I-PER ?/O ...
```

- **KBSET Vanilla** configuration
  - GND until year of birth 1850
  - context year 1789
  - word list includes old orthography

Comparison with the Stanford Named Entity Recognizer

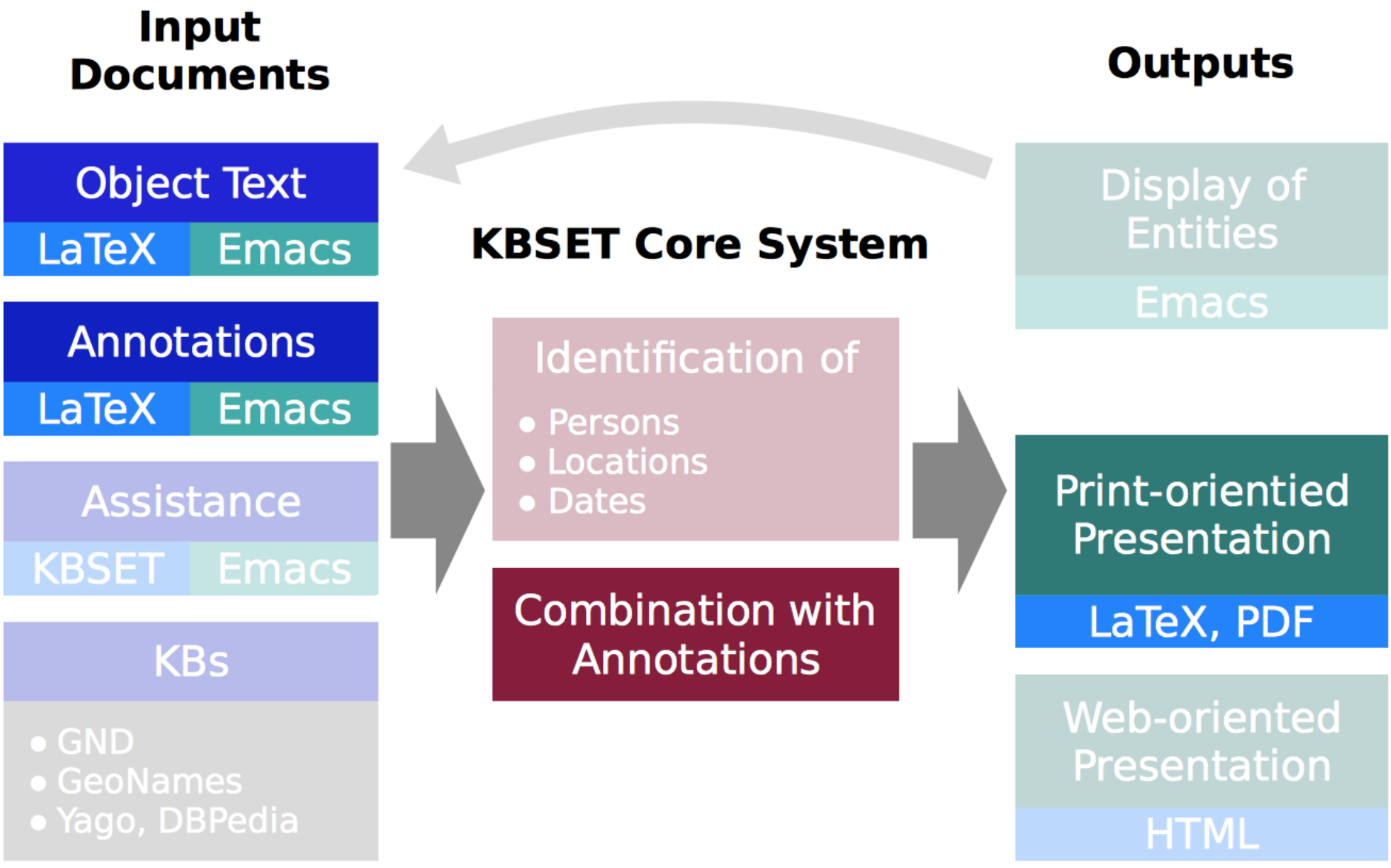
Recognized occurrences of person designators in
Stirner, *Geschichte der Reaction*, Vol. 1, 1852



- Identification incorrect
- Due to old orthography
- Not recognized by KBSET Assisted – hard to identify or not in GND extract

Runtimes: KBSET 25 sec, SNER 20 sec incl. 10 sec classifier loading

Document Combination



LaTeX/ PDF Output

Auszugsweise Übersetzung von Comte: Cours de Philosophie Positive, Bd. 5, S. 583–584.

Heinrich ⟨VIII., England, König⟩ (1491-1547)

Rom

Karl ⟨V., Heiliges Römisches Reich, Kaiser⟩ (1500-1558)

Franz ⟨I., Frankreich, König⟩ (1494-1547)

Auszugsweise Übersetzung von ebd., Bd. 5, S. 585–589.

Österreich

Luther, Martin (1483-1546)

desselben bildete, auch das Seinige bei.

Diese Unterwerfung der geistlichen Gewalt unter die weltliche, diese Grundstörung des ganzen bisherigen Lebenssystems gehörte übrigens dem ganzen weltlichen Europa gemeinsam an, und als Heinrich VIII. von Rom sich lossagte, waren Carl V. und Franz I. nicht weniger emancipirt als er.

Die Entstehung der modernen Reaktion.

An dieser ersten revolutionären Umwandlung haben also die katholischen Völker nicht weniger Theil genommen als die protestantischen.

Automatically generated

- **margin notes for entities**
- **indexes**
- **hyperlinks**
  - within the document
  - to Wikipedia, GND, etc.

, in Oesterreich
und im Grunde
Zeiten Luthers
e Herren, nicht
rotestantischen

r calvinischen
e Folge, indem
terwerfung un-
238] bis dahin,
un die einzige
des drohenden
n sehen mußte.
m katholischen
iche und natur-
llstand, sodann

den Rückschritt bezweckte.

Es braucht übrigens nicht erst bewiesen zu werden, daß diese wachsende Opposition gegen den Fortschritt der menschlichen Entwickelung, fern davon, nur dem modernen, gallicanischen oder spanischen Katholicismus eigen zu sein, auf eine viel gründlichere

External Annotations (Stand-off Markup)

Auszugsweise Übersetzung von Comte: Cours de Philosophie Positive, Bd. 5, S. 583–584.

Heinrich ⟨VIII., England, König⟩ (1491-1547)

Rom

Karl ⟨V., Heiliges Römisches Reich, Kaiser⟩ (1500-1558)

Franz ⟨I., Frankreich, König⟩ (1494-1547)

Auszugsweise Übersetzung von ebd., Bd. 5, S. 585–589.

Diese Unterwerfung der geistlichen Gewalt unter die weltliche, diese Grundstörung des ganzen bisherigen Lebenssystems gehörte übrigens dem ganzen weltlichen Europa gemeinsam an, und als Heinrich VIII. von Rom sich lossagte, waren Carl V. und Franz I. nicht weniger emancipirt als er.

Die Entstehung der modernen Reaktion.

An dieser ersten revolutionären Umwandlung haben also die katholischen Völker nicht weniger Theil genommen als die protes-

emacs

File Edit Options Buffers Tools TeX Development Help

```
Diese Unterwerfung der geistlichen Gewalt unter die
weltliche,
diese Grundstörung des ganzen bisherigen Lebenssystems
gehörte übrigens dem ganzen weltlichen Europa gemeinsam an,
und als Heinrich VIII. von Rom sich lossagte, waren Carl V.
und Franz I. nicht weniger emancipirt als er.

% 410^ Rome, Charles-Quint et François I

\xsecnc{\xuqi Die Entstehung der modernen Reaktion}

% 411v que les peuples catholiques n'aient tout aussi

An dieser ersten revolutionären Umwandlung haben also
die katholischen Völker nicht weniger Theil genommen als die
-U:---  reaction_01_16.tex   77% (??,0)    Git:master  (LaTeX Kbset Fill)-------

\xabout{source}{txt='Diese Unterwerfung der geistlichen Gewalt unter die'}
    {Auszugsweise Übersetzung von \volcite{5}[S.~583--584]{comte:cours}.}

\xabout{source}{txt='An dieser ersten revolutionären Umwandlung haben also'}
    {Auszugsweise Übersetzung von \volcite{5}[S.~585--589]{comte:cours}.}

-U:---  annot_reaction_01.tex   83% (388,0)   Git:master  (LaTeX Fill)----------
```

Some Future Issues on Document Composition

- **Semantics-based** conditions to specify positions to be modified in the object text, e.g. "in the chapters about ..."
- Relating to concepts of **aspect-oriented programming**:

Position	Joint point
Set of positions	Pointcut
Specifier of a set of positions	Pointcut designator
Action to be performed at all positions in a set	Advice
Effecting execution of advices	Weaving

Further Implemented Functionality

- Persons characterized by **function**: "Bishop of Chartres"
- Consideration of **document structure**
- **Keyword extraction**

1. Scholarly Editing
2. Relevant Knowledge Sources
3. KBSET – An Experimental Platform
4. **Coupling Fuzzy and Symbolic Knowledge**
5. Access Predicates
6. Conclusion

Use of Features in the Named Entity Identification of KBSET

Gleim, Johann Wilhelm Ludwig (1719-1803)

Lehrer, Schriftsteller, Sekretär

http://de.wikipedia.org/wiki/Johann_Wilhelm_Ludwig_Gleim

http://d-nb.info/gnd/118717758

Not explicitly blocked	not_explicitly_blocked
	explicitly_specified_in_context(_)
	followed_by_matching_roman_number(_)
	preceded_by_matching_first_names(_)
	explicitly_specified
	preceded_by_matching_first_names_initials(_)
	followed_by_matching_extra_names(_)
	followed_by_matching_extension(_)
	occupation_mentioned(_)
No stop word	no_stopword
No common noun	no_common_substantive
Not commonly used in lowercase	no_common_downcase
No common location name	no_common_geoname
No common first name	no_common_firstname
	already_identified_in_context
Linked to person in context: Sulzer [...]	linked_to_person_in_context(_)
The preferred name is in the text	referenced_by_preferred_name
Linked to more than 50 others	linked_to_many_others(_)
Born 1719, matching context year 1760	born_in_span_before_year_in_context(_, _)
In the German Wikipedia	in_wikipedia_de
Linked to 76 others	linked_to_others(_)

Simple Plausibility Vector Model Currently Used in KBSET

$$\begin{aligned}\text{plausibility}(\text{denotes}(\mathit{WordOccurrence}, \mathit{Entity})) = \langle V_1, \ldots, V_n \rangle \quad \equiv \\ \text{value}(\text{feature}_1, \mathit{WordOccurrence}, \mathit{Entity}) = V_1 \\ \wedge \quad \ldots \\ \wedge \quad \text{value}(\text{feature}_n, \mathit{WordOccurrence}, \mathit{Entity}) = V_n\end{aligned}$$

- Vectors $\langle V_1, \ldots, V_n \rangle$ are compared lexically
- For given $\mathit{WordOccurrence}$ entities are arranged in equivalence classes
  - if the first is a singleton, $\mathit{WordOccurrence}$ is taken as "identified"
- Feature values can depend on
  - previous entity identifications
  - context of $\mathit{WordOccurrence}$
- Vectors $\langle V_1, \ldots, V_n \rangle$ also serve as justifications

1. Scholarly Editing
2. Relevant Knowledge Sources
3. KBSET – An Experimental Platform
4. Coupling Fuzzy and Symbolic Knowledge
5. **Access Predicates**
6. Conclusion

Knowledge Sources have to be Preprocessed for Applications

- Given are source facts

```
rdf_triple(p1, name, 'Sulzer').
rdf_triple(p1, year_of_birth, 1720).
```

- The knowledge is accessed from the application in "directed" ways

```
name_to_year_of_birth(+N, -Y) :-
        name_to_person(+N, -P),
        person_to_year_of_birth(+P, -Y).
```

- It seems useful to precompute indexed access predicates

```
name_to_person(+N, -P)
person_to_year_of_birth(+P, -Y)
```

Tasks to be Automated – with Provers

- **Determine required access predicates** from given queries
- **Rewrite queries** in terms of access predicates
- **Rewrite to subqueries** for different knowledge sources

Definability as Validity and Definientia as Craig Interpolants

- Some second-order entailments can be reduced to first-order entailments:

$$\exists p\, F[p] \models \forall q\, G[q] \quad \text{iff} \quad F[p'] \models G[q'], \qquad p', q' \text{ fresh}$$

- **Definability** of p within F can be expressed as follows:

$$\begin{array}{lll} & \text{There is a } Hx \text{ s.th. } & F \models \forall x\, px \leftrightarrow Hx, & p \text{ not in } Hx \\ \text{iff} & \text{There is a } Ha \text{ s.th. } & F \models pa \leftrightarrow Ha, & p \text{ not in } Ha \\ \text{iff} & \text{There is a } Ha \text{ s.th. } & \exists p\, F \wedge pa \models Ha \models \neg(\exists p\, F \wedge \neg pa), & p \text{ not in } Ha \end{array}$$

a is fresh

- **Craig interpolation** allows construction of **definientia** Ha from proofs
- Generalizations:
  - complex formulas instead of p
  - specific predicates and constants allowed in Ha
  - specific polarity of predicate occurrences in Ha (Lyndon interpolation)

Very Simple Example

$$accessor\_spec \stackrel{\text{def}}{=}$$
$$(\forall pn\ b(p) \rightarrow (person\_name(p,n) \leftrightarrow person\_name\_bf(p,n))) \quad \wedge$$
$$(\forall pn\ b(n) \rightarrow (person\_name(p,n) \leftrightarrow person\_name\_fb(p,n))) \quad \wedge$$
$$(\forall pn\ person\_name(p,n) \rightarrow b(p) \wedge b(n)).$$

$$rewrite_1 \stackrel{\text{def}}{=}$$
$$definiens(person\_name(p,n),$$
$$accessor\_spec \wedge b(n),$$
$$[person\_name\_bf, person\_name\_fb]).$$

$rewrite_1$ expands into a valid implication

$$(\exists b person\_name\ accessor\_spec \wedge b(n) \wedge person\_name(p,n)) \quad \rightarrow$$
$$\neg(\exists b person\_name\ accessor\_spec \wedge b(n) \wedge \neg person\_name(p,n)).$$

Recall: $\exists p\, F \wedge pa \models Ha \models \neg(\exists p\, F \wedge \neg pa)$, p not in Ha

A Craig interpolant for $rewrite_1$ is

$$person\_name\_fb(p,n).$$

Subqueries for Different Knowledge Sources

- Craig interpolation (inductive interpolation [Craig, 1957, Lemma 2]) can compute H_i, each with different restrictions on allowed predicates and constants such that

$$F \models \forall x\, Gx \leftrightarrow H_1 x \vee \ldots \vee H_n x$$

An Example with two Dependent Atoms

$$rewrite_2 \stackrel{\text{def}}{=} definiens(\exists p\, person\_name(p, n) \wedge person\_yob(p, y), accessor\_spec \wedge b(n), [person\_name\_bf, person\_name\_fb, person\_yob\_bf, person\_yob\_fb]).$$

A Craig interpolant of $rewrite_2$ is

$$\exists x\, person\_yob\_bf(x, y) \wedge person\_name\_fb(x, n).$$

An Example with a Referential Constraint

$$person\_spec \stackrel{\text{def}}{=} (\forall p\, person(p) \rightarrow b(p)) \wedge (\forall pn\, person\_name(p, n) \rightarrow person(p)).$$

$$rewrite_3 \stackrel{\text{def}}{=} definiens(\exists p\, person\_name(p, n), person\_spec \wedge accessor\_spec, [person, person\_name\_bf, person\_name\_fb]).$$

A Craig interpolant of $rewrite_3$ is

$$\exists x\, person\_name\_bf(x, n) \wedge person(x).$$

Implementation Framework "ToyElim 2"

- Addressed Issues
  - construction of **complex formalizations**
  - **machine evaluation** of these
  - **reproducible computational tasks** as by-product
- **Prolog-based** system
- Supported core operations for first-order logic
  - **proving**
  - **interpolant computation**
  - **second-order quantifier elimination**
- **Macros**
  - formula labels
  - to specify e.g. `definiens(Q,F,S)`, `is_transitive(P)`
- **LaTeX formula pretty printer**
- Support for **brief syntax**: px
- **TPTP** and **DIMACS** import/export
  - interface to first-order provers and SAT solvers

Prover Used for Interpolation

- CM prover (1992,1997,2015): PTTP/SETHEO/PROTEIN/leanCoP-like
  - model elimination / connection method / **clausal tableaux**
- **Extraction of first-order interpolants**
  - variant of the Smullyan/Fitting method
  - no change of the core prover needed

CM Prover: Performance on the CASC-25 (2015) FOF Problems

Without Equality

Prover	Solved from 150
Vampire 2.6	144
Vampire 4.0	139
iProver	125
E	122
ET	119
CVC4	111
$\bigcup$ CM lean, low-5, std, hd-2	102
CM lean	94
CM low-5	93
iProverModulo	91
CM std	90
leanCoP	85
CM hd-2	85
ePrincess	35
Prover9	29
Muscadet	18
Geo-III	15

With Equality

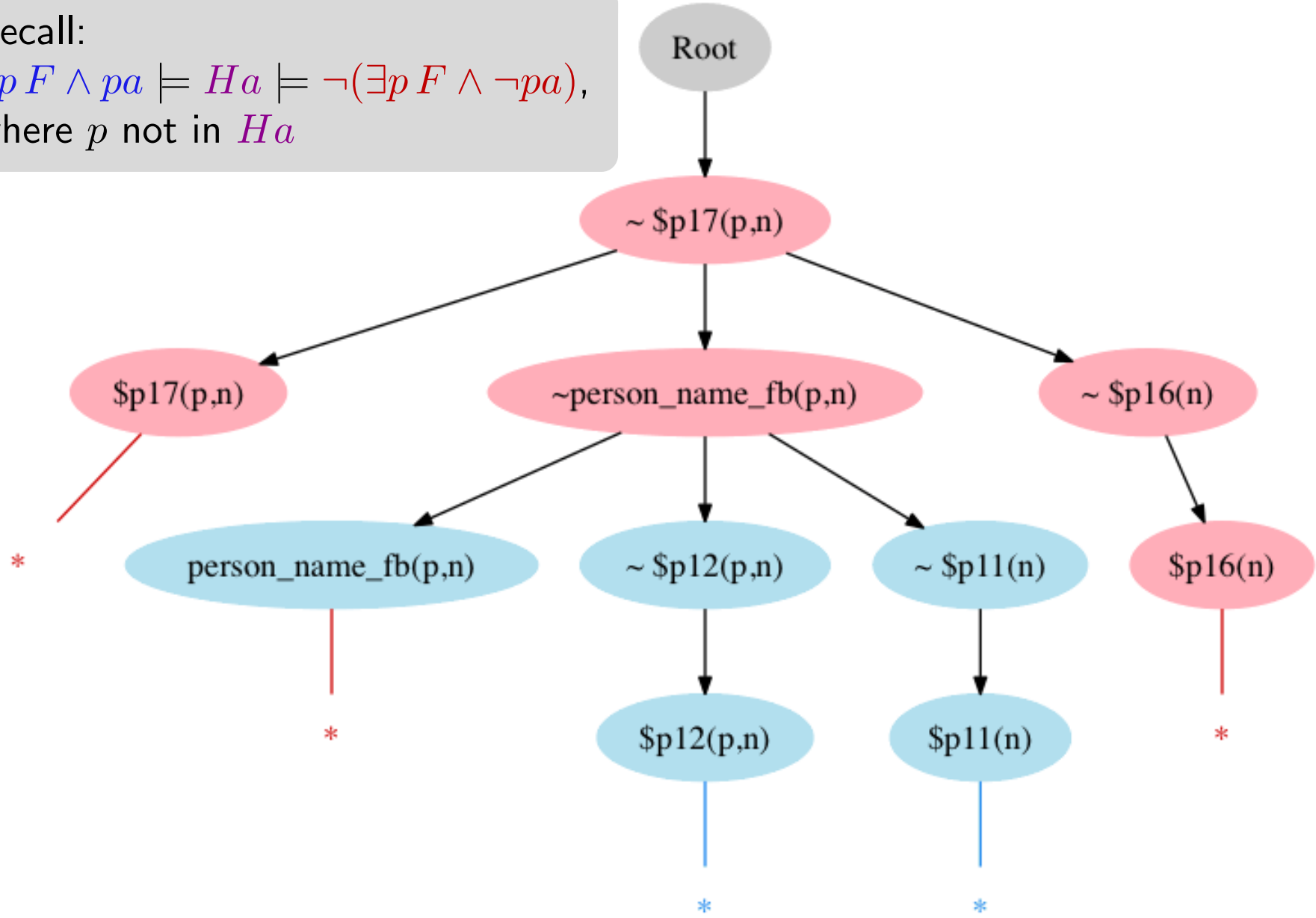
Prover	Solved from 250
Vampire 4.0	241
Vampire 2.6	227
E	194
ET	184
CVC4	146
iProver	97
Prover9	82
ePrincess	78
leanCoP	74
$\bigcup$ CM std, lean, low-5, lem-hd	56
CM std	48
CM lean	46
CM low-5	44
CM lem-hd	42
iProverModulo	36
Geo-III	22
Muscadet	19

CM: 300 sec, 3GB, 3.50GHz, inputs from TPTP 6.3.0, with SWI Prolog

CASC: 300 sec, 32GB, 2.40GHz, axiom preloading and analysis allowed

Clausal Tableau for the Very Simple Example

Recall:
$\exists p\, F \wedge pa \models Ha \models \neg(\exists p\, F \wedge \neg pa)$,
where p not in Ha



Related Works

- Relativized quantifiers to ensure "evaluability"
  [Marx, 2007], [Bárány et al., 2013], [Nash et al., 2010]
- Relativized quantifiers to associate binding patterns
  [Benedikt et al., 2014]
- Comparing interpolants w.r.t. query plan cost estimations
  [Toman and Weddell, 2011, Hudek et al., 2015]

On the ToDo List

- **Rewriting target languages:**
  - what are useful properties for evaluation by proving techniques?
  - what properties of interpolants can be ensured with specific calculi?
- **Auxiliary access predicates** with compound definitions
- **Global "selection conditions"** should be propagated

  e.g. persons born before 1850
- Some **prover control** seems useful:
  - preference of "smaller" proofs
  - preference by ordering on predicate names

1. Scholarly Editing
2. Relevant Knowledge Sources
3. KBSET – An Experimental Platform
4. Coupling Fuzzy and Symbolic Knowledge
5. Access Predicates
6. **Conclusion**

Interesting Aspects from the Viewpoint of Scholarly Editing

- Inclusion of **automated techniques** like named entity identification
- **Embedding and automated use of large external KBs** like GND
- **Combination of KBs with adjustments to achieve precise results**

Focusing on the exceptions,
where automated techniques fail

- Inclusion of **external and generated markup**
- **High-quality presentations** with low cost

Involved Languages/Logics and Methods for them to Develop Further

- **Internal access language**
  - ordered solution sets
  - support for justifications
  - automated generation of access predicates
    $\Rightarrow$ interpolation
- **Assistance language to adjust entity identification**
  - focus on exceptions
    $\Rightarrow$ non-monotonic reasoning
- **Assistance language to control document combination**
  - specifying sets of text positions
  - specifying modifications to be performed at these

  $\Rightarrow$ similarites to aspect-oriented programming
- **Semantics-based modularization** by forgetting about subvocabularies
  $\Rightarrow$ second-order quantifier elimination

References

[Benedikt et al., 2014] Benedikt, M., ten Cate, B., and Tsamoura, E. (2014).

Generating low-cost plans from proofs.

In *PODS'14 – Proc. 33rd ACM SIGMOD-SIGACT-SIGART Symp. on Principles of Database Systems*, pages 200–211. ACM.

[Bárány et al., 2013] Bárány, V., Benedikt, M., and ten Cate, B. (2013).

Rewriting guarded negation queries.

In *Mathematical Foundations of Computer Science 2013 (MFCS 2013)*, volume 8087 of *LNCS*, pages 98–110. Springer.

[Craig, 1957] Craig, W. (1957).

Three uses of the Herbrand-Gentzen theorem in relating model theory and proof theory.

JSL, 22(3):269–285.

[Finkel et al., 2005] Finkel, J. R., Grenager, T., and Manning, C. (2005).

Incorporating non-local information into information extraction systems by Gibbs sampling.

In *Proc. 43nd Ann. Meeting of the Association for Computational Linguistics (ACL 2005)*, pages 363–370. ACL.

[Hudek et al., 2015] Hudek, A., Toman, D., and Wedell, G. (2015).
On enumerating query plans using analytic tableau.
In *TABLEAUX 2015*, volume 9323 of *LNCS (LNAI)*, pages 339–354. Springer.

[Marx, 2007] Marx, M. (2007).
Queries determined by views: Pack your views.
In *PODS '07*, pages 23–30. ACM.

[Nash et al., 2010] Nash, A., Segoufin, L., and Vianu, V. (2010).
Views and queries: Determinacy and rewriting.
TODS, 35(3).

[Toman and Weddell, 2011] Toman, D. and Weddell, G. (2011).
Fundamentals of Physical Design and Query Compilation.
Synthesis Lectures on Data Management. Morgan and Claypool.