Computer Support for the Analysis and Improvement of the Readability of IT-related Texts

Matthias Holdorf, 23.05.2016, Munich

Software Engineering for Business Information Systems (sebis)
Department of Informatics
Technische Universität München, Germany

wwwmatthes.in.tum.de

Administrative Setup

- **Chair**: Software Engineering for Business Information Systems
- **Company**: QAWare GmbH

- **Title**: Computer Support for the Analysis and Improvement of the Readability of IT-related Texts
- **Advisor**: Bernhard Waltl (b.waltl@tum.de)
  Andreas Zitzelsberger (andreas.zitzelsberger@qaware.de)

- **Author**: Matthias Holdorf (matthias.holdorf@gmail.com)
- **Start**: 15. May 2016
- **Submission:** 15. November 2016

Motivation

„Die Probe der Güte ist, dass der Leser nicht zurückzulesen hat.“

„The sample of kindness is, that the reader does not have to read back.“

– Jean Paul

Research Approach: Design Science Approach [1], [2]

sebis

(1) Identify Business Problems

- Guided and expert interviews
- Show mock-ups early

(2) Develop a Design for an Artefact

- Create evaluation criteria

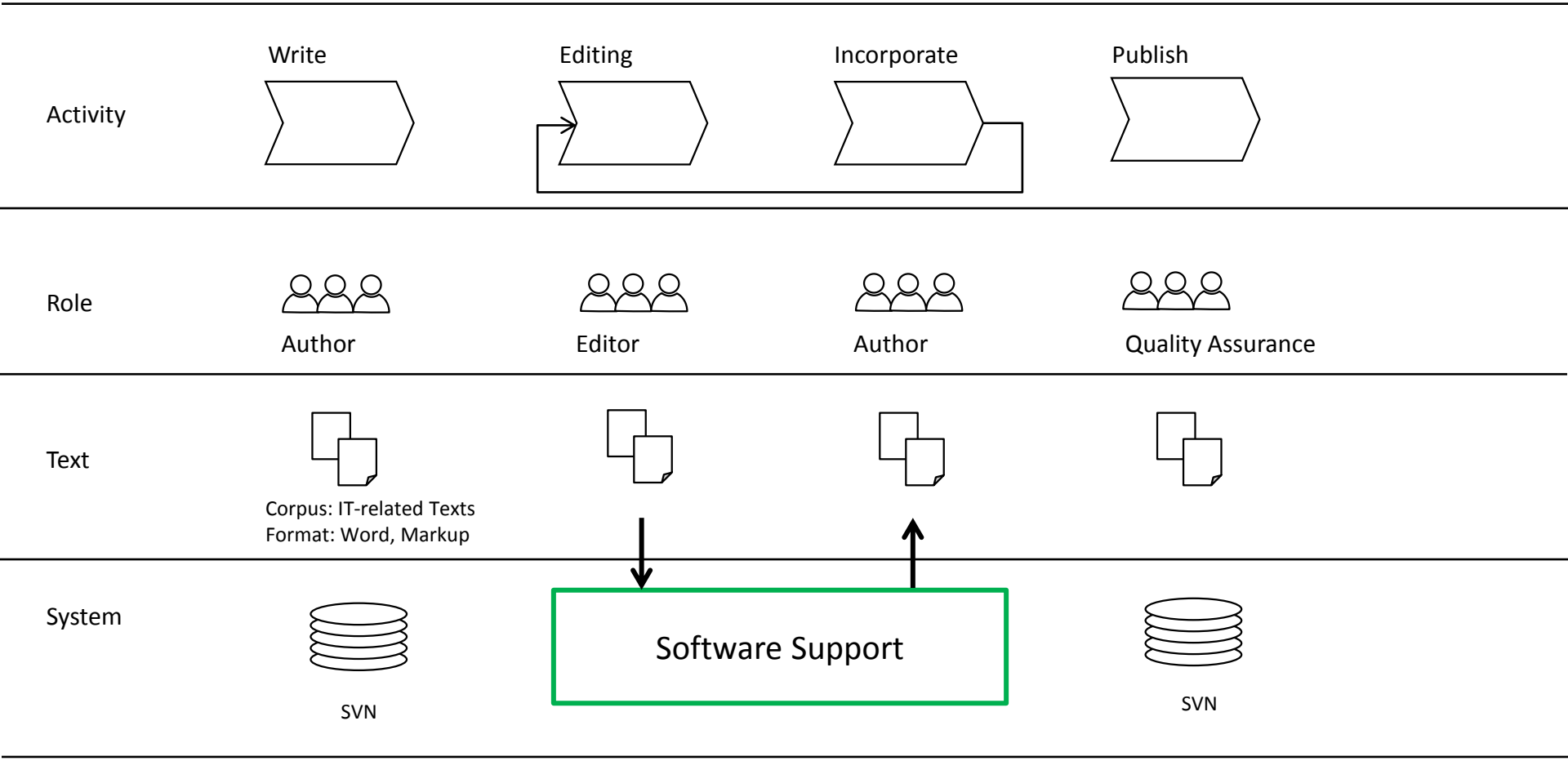
(3) Design and Implementation

- Strong focus on executable artefact
- Demonstrate artefact

(4) Evaluation

- Guided and expert interviews

Process Model of Editing a Text

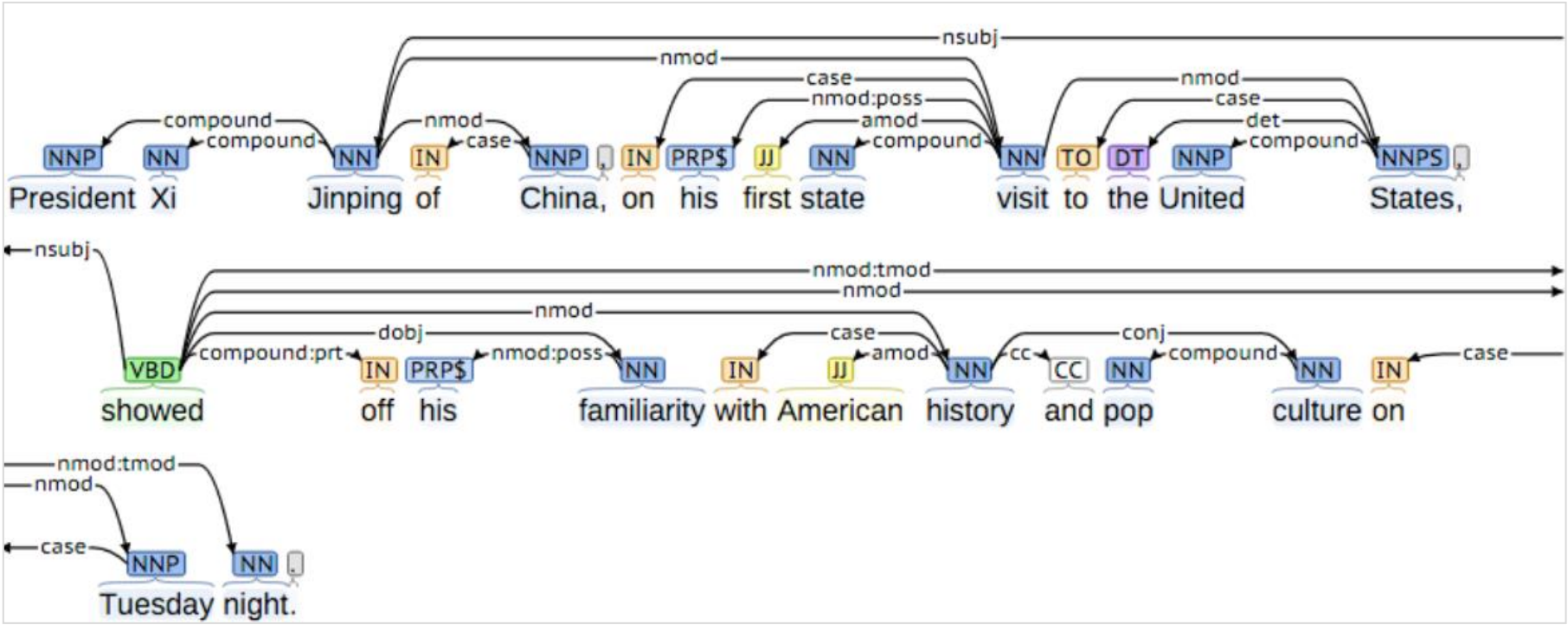


Readability Formulas on Document Level [3]

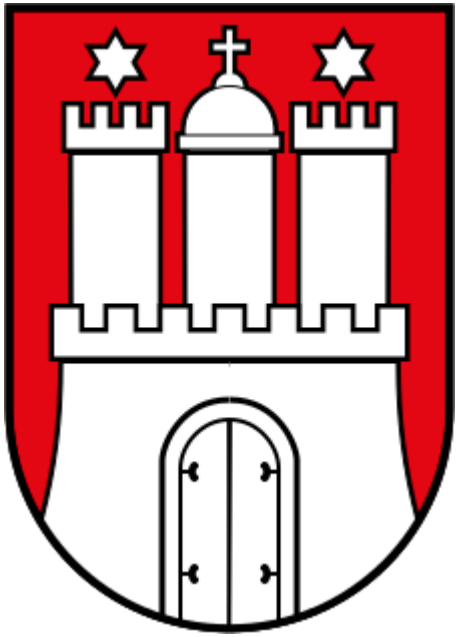
$$FRE = 206{,}835 - 1{,}015\left(\frac{\text{Total Words}}{\text{Total Sentences}}\right) - 84.6\left(\frac{\text{Total Syllables}}{\text{Total Words}}\right)$$

Score	Readability	Example
0–30	Very difficult	Academic
30–50	difficult	
50–60	Fairly difficult	
60–70	Mediocre	13–15-year old Students
70–80	Fairly easy	
80–90	Easy	
90–100	Very easy	11-year old Students

Readability Formulas on Sentence Level I [4]



Readability Formulas on Sentence Level II [5]

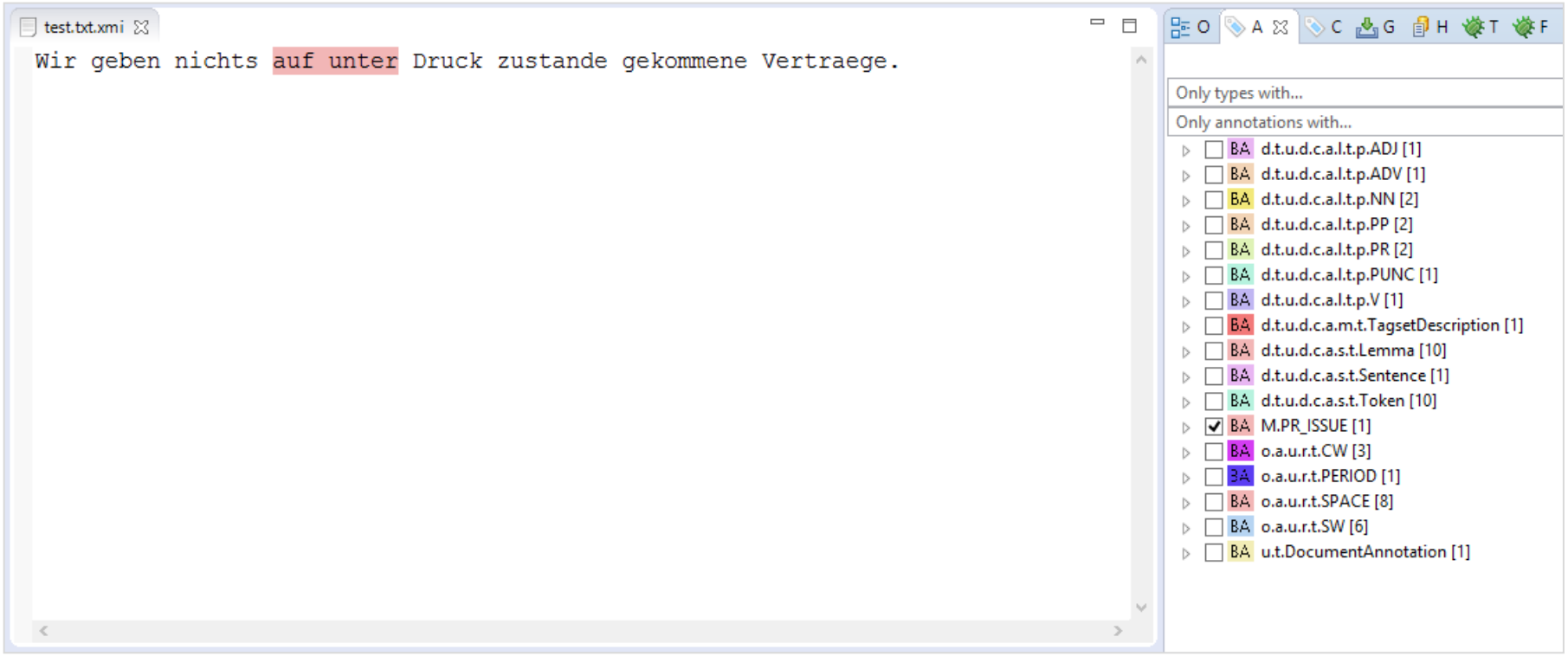


Hamburger Verständlichkeitskonzept

Readability Formulas on Sentence Level III [6]

Apache UIMA Ruta

A language for rule-based text annotations.



Sample of edited Sentences (Mock-up)

A possible representation of improvements of the readability is presented during the interviews, ***after*** the interviewee gave us feedback on how such improvement may be incorporate and visualized in a document.

A.3 Example of edited Sentences

Wir geben nichts auf unter Druck zustande gekommene Verträge.

Kommentar [M2]: Zwei aufeinanderfolgende Präpositionen sind störend zu vermeiden.

Inzwischen stellen regionale Bezüge bzw. ein entsprechend zu lokalen Zugehörigkeiten und Erfahrungen getönter Hintergrund sowohl im Bereich [..] auch in der neuen Bundesrepublik Deutschland eine zentrale Dimension dar.

Kommentar [M3]: Das Verb „darstellen" ist durch mehr als 6 Wörter getrennt, das führt zu einer schwer zu verstehenden Verständlichkeit.

Ein schleichender, von den Nutzern typischerweise durch Aussagen wie „Das ist so langsam" oder „Die Zahlen taugen nichts" kommunizierter Qualitätsverlust.

Kommentar [M4]: Zwischen dem Artikel „Ein" und das Substantiv „Qualitätsverlust" sind mehr als 3 vorangestellte Attribute.

Folglich stiegen die Hoffnungen, dass die Bewältigung der kommenden Herausforderungen und die Anpassung der Wirtschaftsordnung an die veränderten Rahmenbedingungen des globalen Wettbewerbs gelingen würden.

Kommentar [M5]: Dieser Satz enthält 6 abstrakte Substantive die auf -ung, -keit oder -ung enden. Ein Satz sollte maximal 2 solcher Substantive enthalten.

Die Versuche der CDU, einen Keil zwischen SPD und FDP zu treiben […] hat der FDP-Vorsitzende scharf verurteilt.

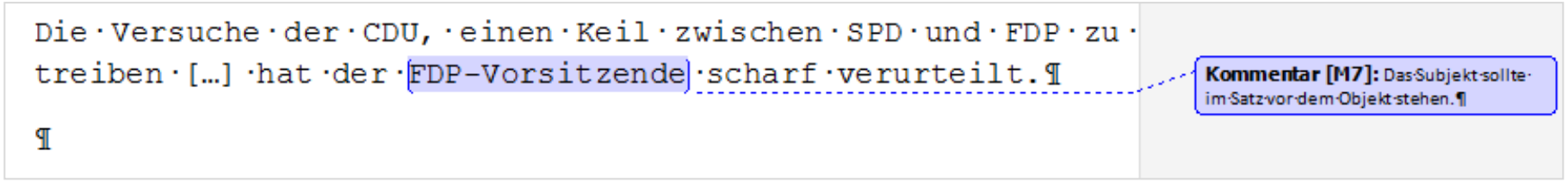
Kommentar [M6]: Das Subjekt sollte im Satz vor dem Objekt stehen.

Bei diesen Verfahren seien ausnahmslos die Befürworter eines Verbotes des Zugverkehrs unterlegen.

Kommentar [M7]: Dieser Satz enthält doppelte Verneinung. Der Durchschnittsmensch benötigt 48% mehr Zeit eine verneinende Aussage zu verstehen.

Dem Fachausschuss sollte bis zum 18. Juni ein Vorschlag für ein Stufenkonzept zum Aufbau einer Notrufzentrale einschließlich der hierfür erforderlichen Zeitspanne unterbreitet werden.

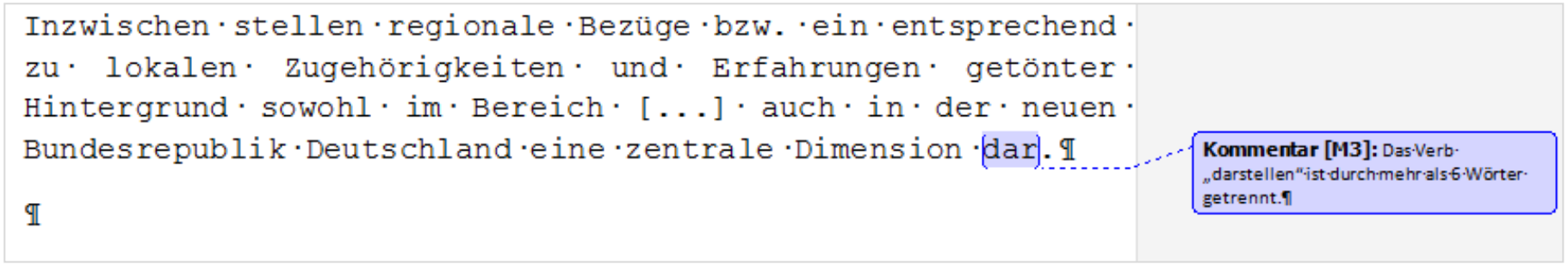
Kommentar [M8]: Bei einer Aufzählung sollte die unmittelbar zweite Hälfte des Verbums nach dem ersten Posten eingeschoben werden.



Die·Versuche·der·CDU,·einen·Keil·zwischen·SPD·und·FDP·zu·treiben·[…]·hat·der·FDP-Vorsitzende·scharf·verurteilt.¶

¶

Kommentar [M7]: Das·Subjekt·sollte·im·Satz·vor·dem·Objekt·stehen.¶

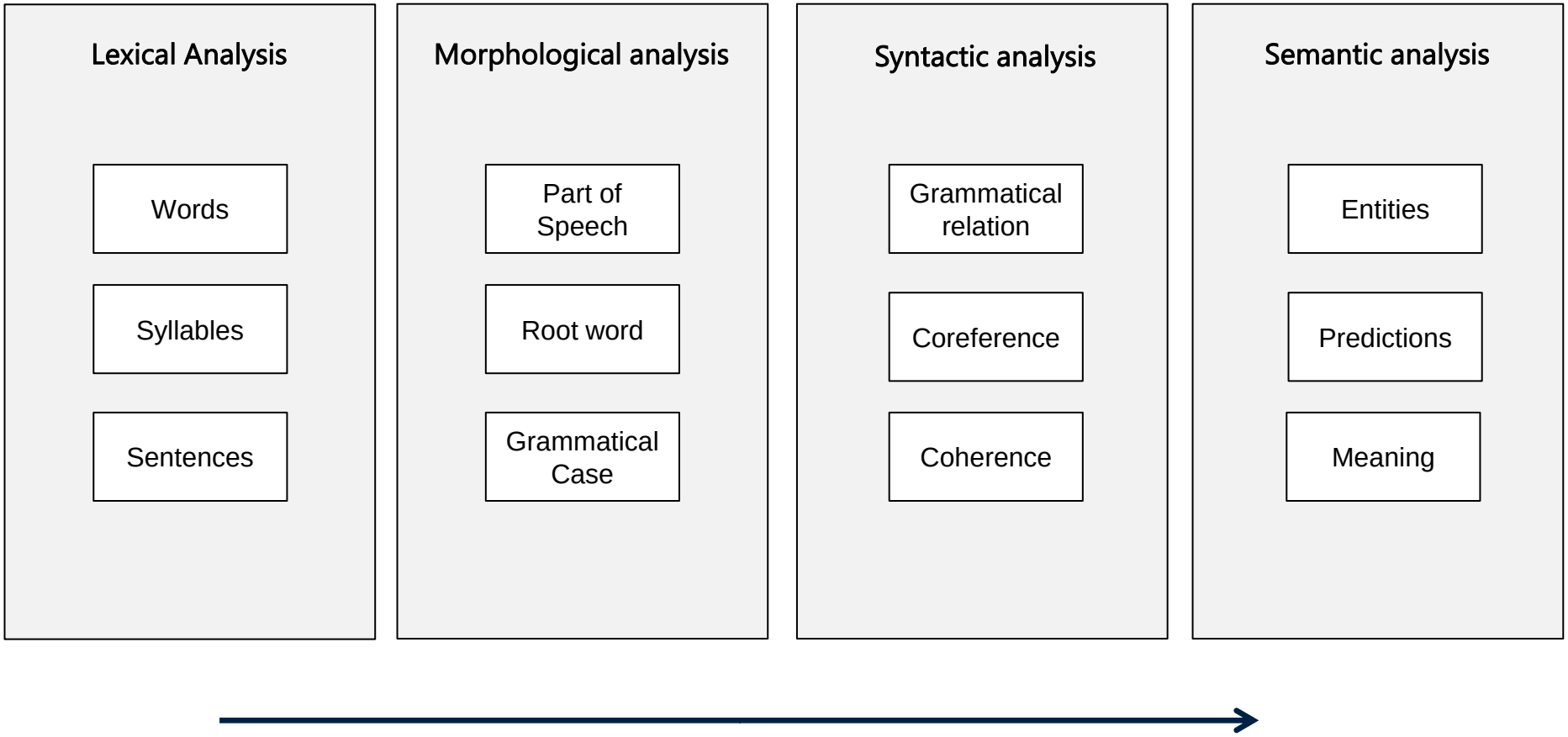


Inzwischen·stellen·regionale·Bezüge·bzw.·ein·entsprechend·zu·lokalen·Zugehörigkeiten·und·Erfahrungen·getönter·Hintergrund·sowohl·im·Bereich·[...]·auch·in·der·neuen·Bundesrepublik·Deutschland·eine·zentrale·Dimension·dar.¶

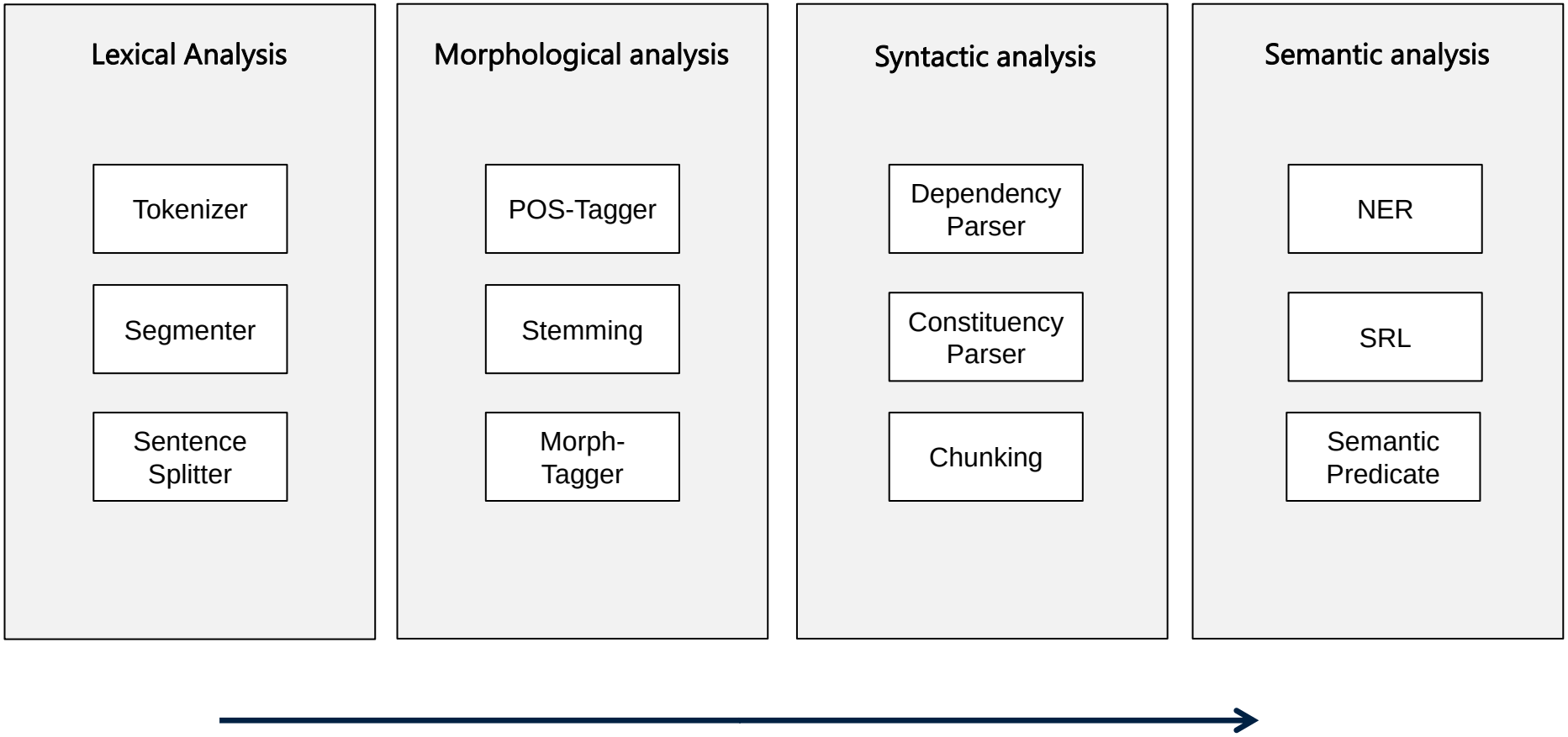
¶

Kommentar [M3]: Das·Verb·„darstellen"·ist·durch·mehr·als·6·Wörter·getrennt.¶

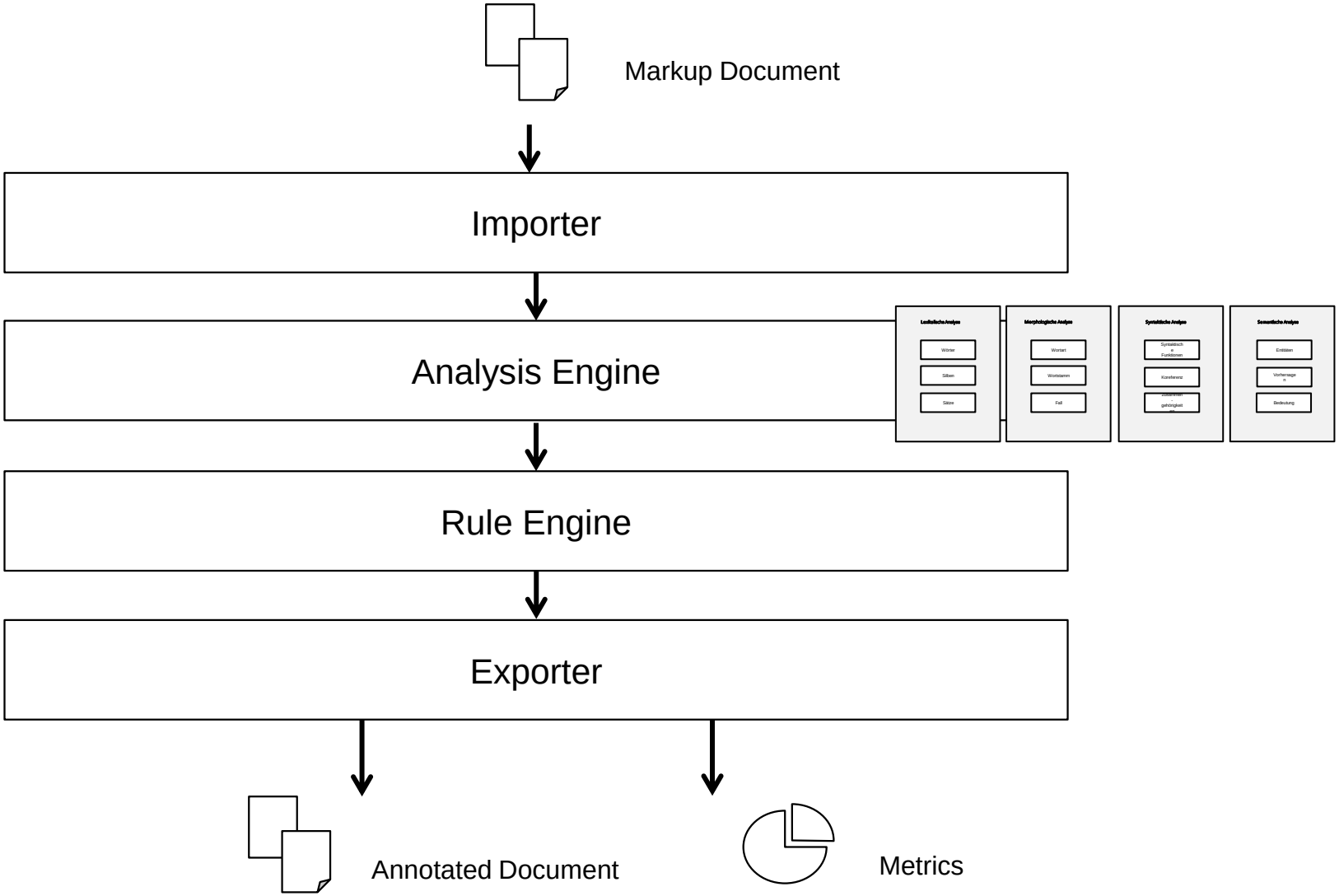
Professional Approach I



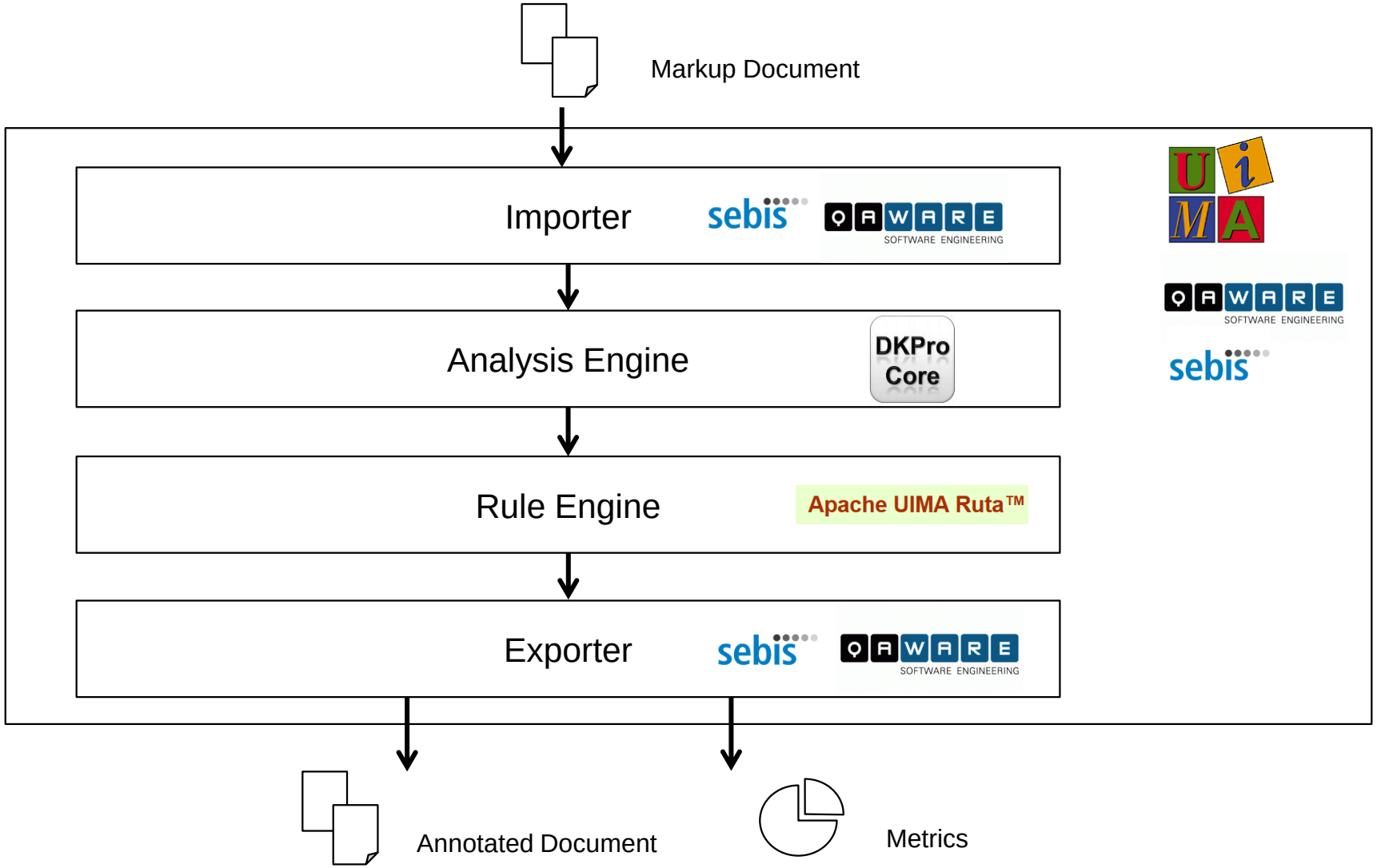
Technical Approach I



Professional Approach II



Technical Approach II



Challenge, Research Questions and Contributions

sebis

How does the **process model** of a document being written, edited and published in an IT-Company looks like?

Which **actor** has which **problems** at a certain **state** in the **process model**?

How can we **improve the readability** of IT-related text? What **readability formulas** and **pattern** exist for the German language?

How can we make the **improvement available to the authors**?

What are the **functional** and **non-functional requirements (evaluation criteria)** of a software to support the analysis and the improvement of the Readability of IT-Related Text?

How does a **prototypical implementation** enabling the analysis and improvement of the Readability of IT-related text look like? How can it be **integrated in the workflow** of an IT-Company?

Project Planning and Milestones

1. Preparatory Phase [01.04 – 29.04]

- Literature research and demarcation
- Deeper familiarization with NLP
- Investigation of NLP architecture and libraries
- Define first structure of paper

2. Developing theoretical foundations [02.05 – 24.06]

- Identify business problems (expert interviews)
- Literature research
- Develop a concept
- Formulation in the thesis paper

3. Implementation [27.06 – 23.09]

- Implementation of the architecture
- Implementation of readability formula
- Demonstration of artefact
- Integration of the prototype

4. Evaluation of the prototype [26.09 – 14.10]

- Evaluation of the prototype (expert interviews)
- Adjustments of the prototype

5. Evaluation and correction phase [14.10– 15.11]

- Formulation in the thesis paper
- Correction of the thesis

Matthias Holdorf
B.Sc. Business Informatics

Technische Universität München
Department of Informatics
Chair of Software Engineering for
Business Information Systems

Boltzmannstraße 3
85748 Garching bei München

Tel. +49 152 534 490 65
E-Mail matthias.holdorf@gmail.com

wwwmatthes.in.tum.de

Bibliography

[1] Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science ininformation systems research. *MIS Quarterly, 28*(1), 75–105.

[2] Hevner, A. (2015): Robust Processes of Design Science Research. https://www.youtube.com/watch?v=gdCYH_a4hzY

[3] Flesch, R. (1948): A New Readability Yardstick. In: *Journal of Applied Psychology* 32(3), 221–233.

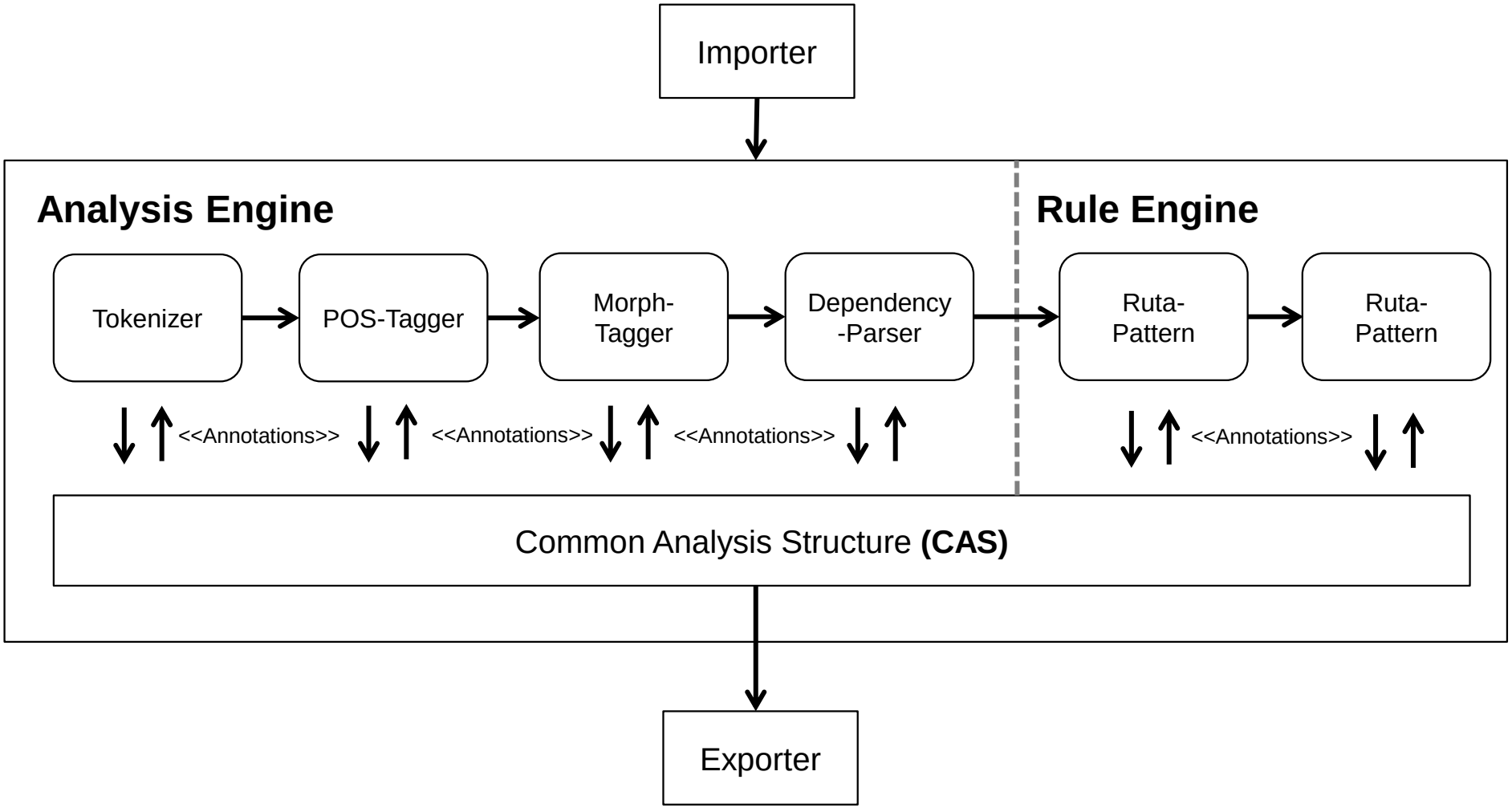
[4] Stanford CoreNLP (2016): A suite of core NLP tools. http://stanfordnlp.github.io/CoreNLP/#about

[5] Schneider, W. (2001): Deutsch für Profis, Wege zu gutem Stil. 23. Aufl., Wilhelm Goldmann Verlag, München, 2001.

[6] Apache UIMA Ruta (2016): Rule-based Text Annotation. https://uima.apache.org/ruta.html

Backup

Prototype Architecture I



Prototype Architecture II