

Integration of graph editor CISGraph for local grammars into WiTTFind

Wittgenstein Summer School 2014
Center for Information and Language Processing
University of Munich

Yuliya Kalasouskaya, B.A.
Advisor: Dr. Maximilian Hadersbeck

06.06.2014

Outline

1. Introduction
2. Local grammars in CISGraph
3. Unix and CISGraph
4. CISGraph in CISWeb 2.0
5. Google Web Toolkit
6. Client-server communication
7. HTML5 Storage
8. Graphical user interface
9. Conclusion
10. Outlook
11. References/List of figures

Introduction

- Wittgenstein Scholarship and Wittgenstein in Co-Text
- WiTTFind is a finder application for searching in texts of Ludwig Wittgenstein



Introduction

Main topics in this bachelor thesis :

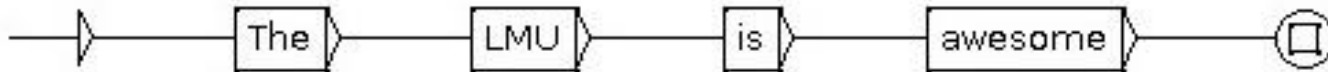
- Extend WiTTFind application with CISGraph
- Perform search queries based on local grammars

Outline

1. Introduction
2. Local grammars in CISGraph
3. Unitex and CISGraph
4. CISGraph in CISWeb 2.0
5. Google Web Toolkit
6. Client-server communication
7. HTML5 Storage
8. Graphical user interface
9. Conclusion
10. Outlook
11. References/List of figures

Local grammars in CISGraph

- Language is a set of sentences with certain syntactical and morphological rules
- Local grammars can be represented as regular expressions, as transformation rules or as a graph:



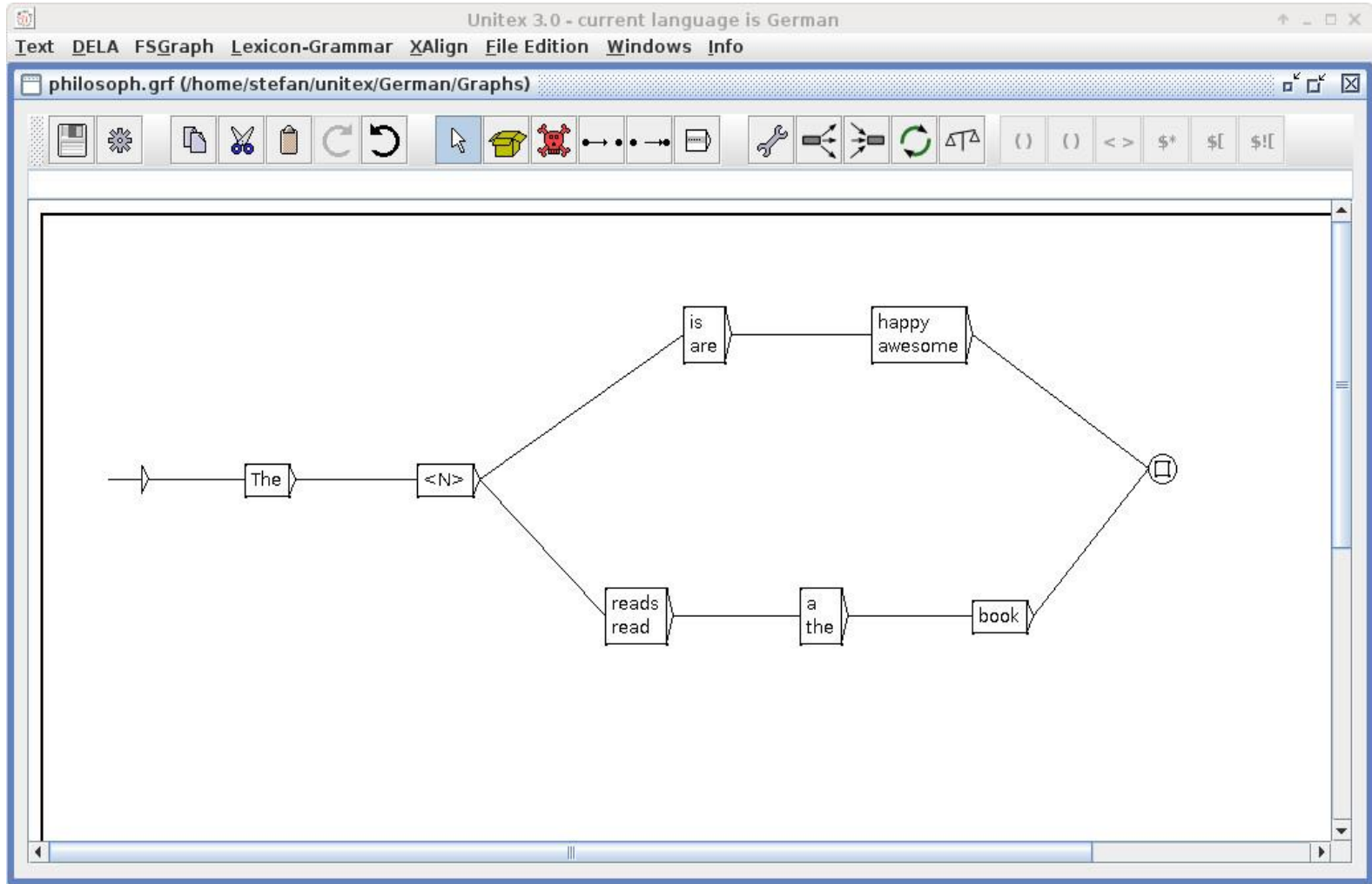
Outline

1. Introduction
2. Local grammars in CISGraph
3. **Unitex and CISGraph**
4. CISGraph in CISWeb 2.0
5. Google Web Toolkit
6. Client-server communication
7. HTML5 Storage
8. Graphical user interface
9. Conclusion
10. Outlook
11. References

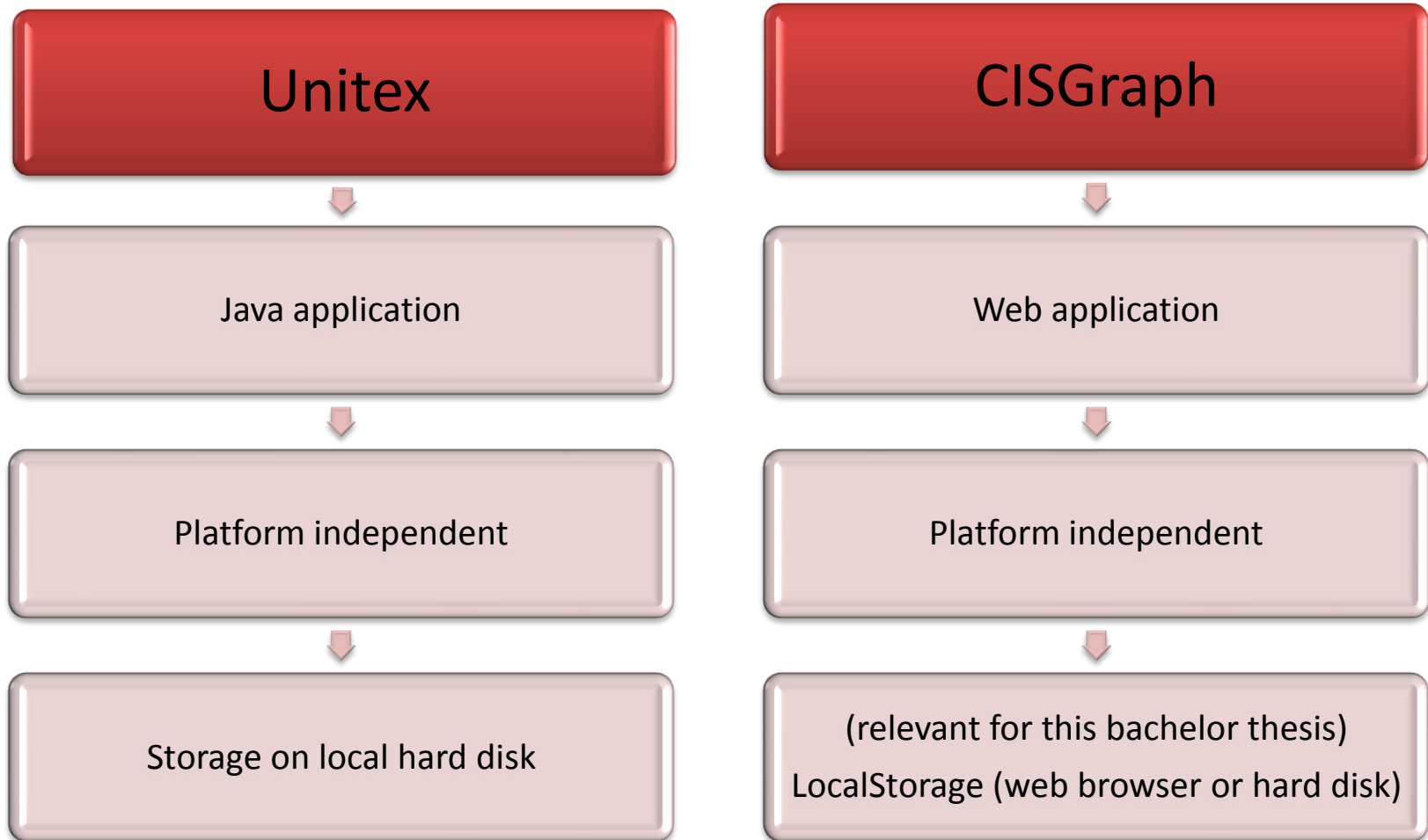
Unitex

- Open-source corpus tool
- Gaspard-Monge Institute of Université de Marne la Vallée
- Fields of Unitex :
 - Use of electronic dictionaries
 - Generating of local grammars
 - Usage of regular expressions for searching

Unitex



Unitex und CISGraph



Outline

1. Introduction
2. Local grammars in CISGraph
3. Unitex and CISGraph
4. **CISGraph in CISWeb 2.0**
5. Google Web Toolkit
6. Client-server communication
7. HTML5 Storage
8. Graphical user interface
9. Conclusion
10. Outlook
11. References/List of figures

CISGraph in CISWeb 2.0

- Shuangjaio Cao, 2012 with a help of Patrick Seebauer
- CISWeb2.0 Tools:



Outline

1. Introduction
2. Local grammars in CISGraph
3. Unitex and CISGraph
4. CISGraph in CISWeb 2.0
5. **Google Web Toolkit**
6. Client-server communication
7. HTML5 Storage
8. Graphical user interface
9. Conclusion
10. Outlook
11. References/List of figures

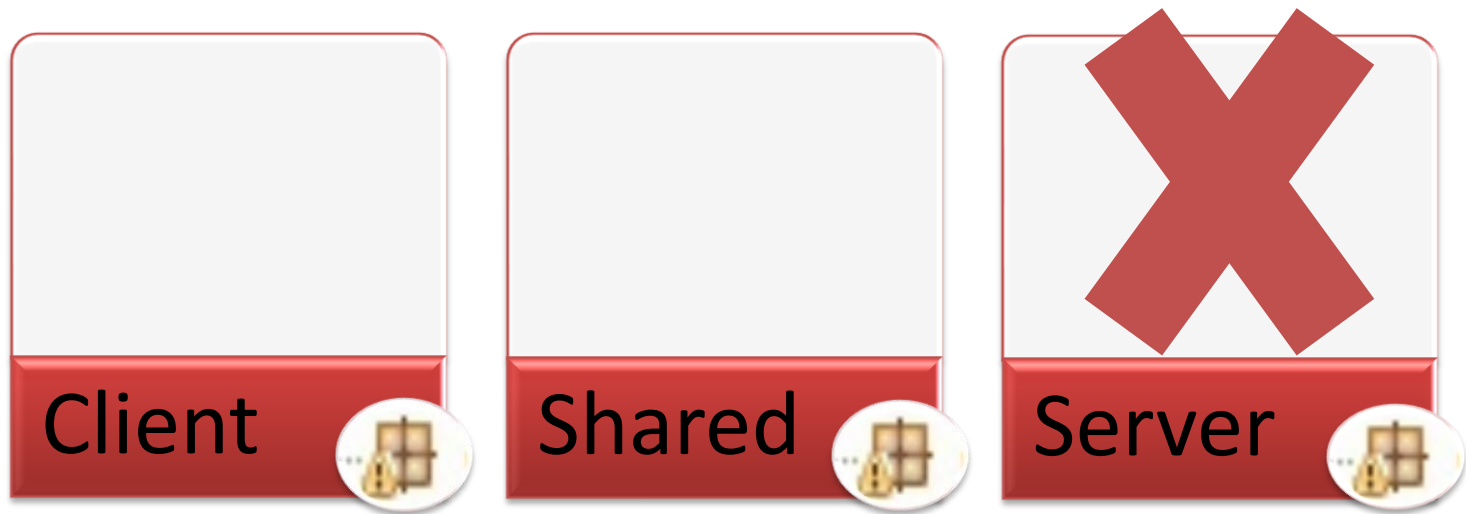
Google Web Toolkit (GWT)

- Released 2006
- Its goal is to enable productive development of high-performance web applications
- Java to JavaScript compiler



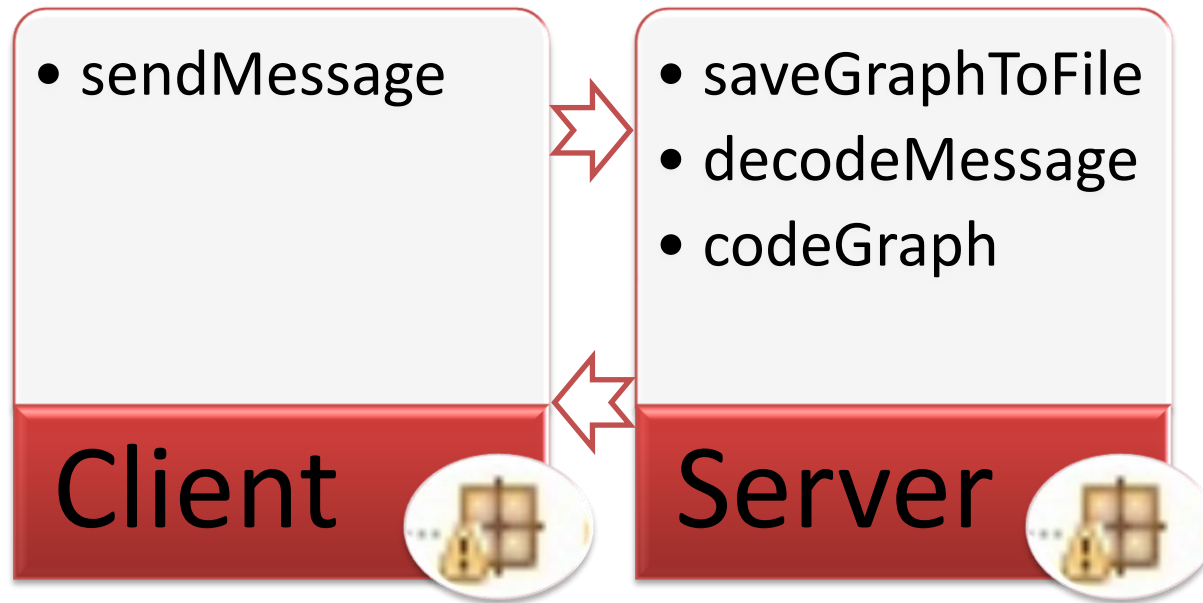
Google Web Toolkit

Former package structure:



Google Web Toolkit (GWT)

Former implementation:



Google Web Toolkit (GWT)

New implementation topics:

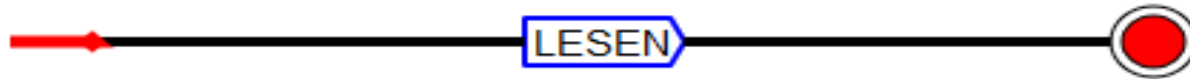
- Implementation of GWT project without built-in GWT server
- Client-side data management
- Implementation of new components
- CISGraph integration into other projects

Outline

1. Introduction
2. Local grammars in CISGraph
3. Unitex and CISGraph
4. CISGraph in CISWeb 2.0
5. Google Web Toolkit
- 6. Client-server communication**
7. HTML5 Storage
8. Graphical user interface
9. Conclusion
10. Outlook
11. References/List of figures

Client-server communication

- JSON data format
- Example of JSON object:

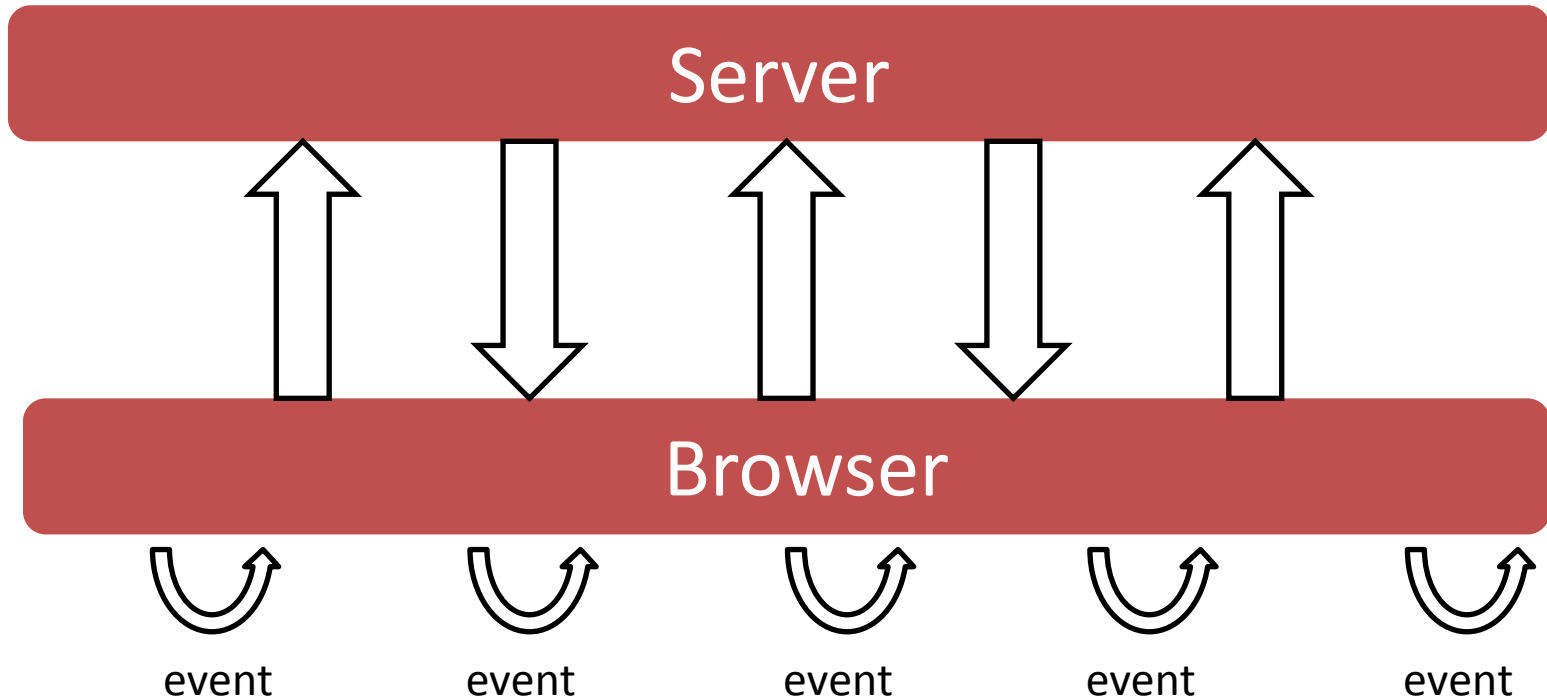


- ```
{"Nodes":{"0":{"type":"START", "x":215, "y":65, "data":{}}, "1":{"type":"FINAL", "x":600, "y":50, "data":{}}, "2":{"type":"MATCH", "x":399, "y":60, "data":"LESEN"}}, "Links":[{"from":0, "to":2}, {"from":2, "to":1}]}
```

JSON object



# Client-server communication



# Client-server communication

WITFind  WITFind-Suche

318 Treffer gefunden

(Ts-213,i-r[7]) [Faksimile](#), [Wittgenstein Source Normalized](#), [Wittgenstein Pundit](#)  
6)Man sagt ein Wort **verstehen** heißt, wissen, wie es gebraucht wird.

6) Man sagt: ein Wort verstehen heisst, wissen, wie es gebraucht wird.  
Was heisst es, das zu wissen? Dieses Wissen haben wir sozusagen im Vor-  
rat.(S.22) BEDEUTUNG ↓

(Ts-213,ä-r[14]) [Faksimile](#), [Wittgenstein Source Normalized](#), [Wittgenstein Pundit](#)  
35)Ein Wort **verstehen** = es anwenden können.

35) Ein Wort verstehen = es anwenden können. Eine Sprache  
verstehen: Einen Kalkül beherrschen. (S.143)  
AUGENBLICKLICHES VERSTEHEN etc. ↓

(Ts-213,ä-r[14]) [Faksimile](#), [Wittgenstein Source Normalized](#), [Wittgenstein Pundit](#)  
Eine Sprache  
**verstehen**: Einen Kalkül beherrschen.

35) Ein Wort verstehen = es anwenden können. Eine Sprache  
verstehen: Einen Kalkül beherrschen. (S.143)  
AUGENBLICKLICHES VERSTEHEN etc. ↓

# Outline

1. Introduction
2. Local grammars in CISGraph
3. Unitex and CISGraph
4. CISGraph in CISWeb 2.0
5. Google Web Toolkit
6. Client-server communication
7. **HTML5 Storage**
8. Graphical user interface
9. Conclusion
10. Outlook
11. References/List of figures

# HTML5 Storage

- Data storage in web browser
- Storage Types:
  - LocalStorage
  - SessionStorage
- Advantages of LocalStorage:
  - Persistent storage of data (browser restart)
  - Prevent work loss from network disconnects
  - Load cached data on startup
  - Reduce network traffic

# Outline

1. Introduction
2. Local grammars in CISGraph
3. Unitex and CISGraph
4. CISGraph in CISWeb 2.0
5. Google Web Toolkit
6. Client-server communication
7. HTML5 Storage
8. Graphical user interface
9. Conclusion
10. Outlook
11. References/List of figures



# Graphical user interface

CIS

CENTRUM FÜR INFORMATIONS UND SPRACHVERARBEITUNG  
WITTGENSTEIN ARCHIVES UNIVERSITY OF BERGEN

WAB

Regelbasiertes Finden   Semantisches Finden   Graphischer Editor   Geheimschriftübersetzer   Statistische Suche   Hilfe

there is a cool graph

New components

You can select and move boxes by clicking on them.

# Graphical user interface

CIS

CENTRUM FÜR INFORMATIONS UND SPRACHVERARBEITUNG  
WITTGENSTEIN ARCHIVES UNIVERSITY OF BERGEN

WAB

Regelbasiertes Finden Semantisches Finden Graphischer Editor Geheimschriftübersetzer Statistische Suche Hilfe

there

there is a cool graph

SAVE

Create a graph and save it in web browser

Einen Graphnamen eingeben!

graph

OK

# Graphical user interface

CIS

CENTRUM FÜR INFORMATIONEN UND SPRACHVERARBEITUNG  
WITTGENSTEIN ARCHIVES UNIVERSITY OF BERGEN

WAB

Regelbasiertes Finden Semantisches Finden Graphischer Editor Geheimschriftübersetzer Statistische Suche Hilfe

there is a cool graph

DOWNLOAD

Download the created graph to hard disk if necessary

Download (2)

Alle Downloads anzeigen...

# Graphical user interface

The screenshot shows a web application interface with a red header bar. On the left, the logo 'CIS' is displayed, and on the right, 'WAB' is displayed. The header text reads 'CENTRUM FÜR INFORMATIONS UND SPRACHVERARBEITUNG WITTMENSTEIN ARCHIVES UNIVERSITY OF BERGEN'. Below the header, a navigation menu contains the following items: 'Regelbasiertes Finden', 'Semantisches Finden', 'Graphischer Editor', 'Geheimschriftübersetzer', 'Statistische Suche', and 'Hilfe'. The main content area features a search bar at the top left and a vertical toolbar on the left with icons for search, save, folder, upload, and other functions. The central workspace displays a graph with nodes labeled 'there', 'is', 'a', 'cool', and 'graph' connected by lines. A red arrow points to the 'there' node, and a red circle highlights the 'graph' node. A large red button with white text reads 'Upload a graph from hard disk into web browser'. Below this button, a dialog box contains two buttons: 'Dateien auswählen' and 'Datei hochladen', with the text 'loadGraph' positioned between them.

# Graphical user interface

CIS

CENTRUM FÜR INFORMATIONS UND SPRACHVERARBEITUNG  
WITTGENSTEIN ARCHIVES UNIVERSITY OF BERGEN

WAB

Regelbasiertes Finden Semantisches Finden Graphischer Editor Geheimschriftübersetzer Statistische Suche Hilfe

there is a cool graph

Geladene Graphen

- [graph](#)
- [loadGraph](#)
- [nom\\_phrase](#)
- [read](#)
- [verb\\_phrase](#)

Select and load an uploaded graph into graphical interface

# Outline

1. Introduction
2. Local grammars in CISGraph
3. Unitex and CISGraph
4. CISGraph in CISWeb 2.0
5. Google Web Toolkit
6. Client-server communication
7. HTML5 Storage
8. Graphical user interface
9. **Conclusion**
10. Outlook
11. References/List of figures

# Conclusion

- CISGraph background
- Integration into WiTTFind
- Removal of GWT server
- Client-side data management and data storage
- HTML5 Storage technology

# Outline

1. Introduction
2. Local grammars in CISGraph
3. Unitex and CISGraph
4. CISGraph in CISWeb 2.0
5. Google Web Toolkit
6. Client-server communication
7. HTML5 Storage
8. Graphical user interface
9. Conclusion
- 10. Outlook**
11. References/List of figures



# Outlook

- More complex search queries
- Change of file name during the storage on hard disk
- Display screenshot of current graph in search result window
- Subdirectories for subgraphs in LocalStorage
- Database of defined graphs

# References

- **Cao, Shuangjiao** (2012): Entwicklung eines Graphischen Editors zur Erstellung von Lokalen Grammatiken innerhalb des Web 2.0 ProjektsCisWeb.
- **Carstensen, Kai-Uwe** (2001): *Computerlinguistik und Sprachtechnologie : eine Einführung*. Heidelberg: Spektrum Akademischer Verlag
- **Hadersbeck, Max, Pichler, Alois, Fink, Florian, Gjesdal, Øyvind Liland** (2014): Wittgenstein's Nachlass: WiTTFind and Wittgenstein Advanced Search Tools (WAST). In: *Proceedings of the First International Conference on Digital Access to Textual Cultural Heritage* (S. 91–96). New York, NY, USA: ACM Online verfügbar unter: URL: <http://doi.acm.org/10.1145/2595188.2595202>
- **Nagel, Sebastian** (2008): *Lokale Grammatiken zur Beschreibung von lokativen Sätzen und ihre Anwendung im Information Retrieval*. Ludwig-Maximilians-Universität München Online verfügbar unter: URL: <http://nbn-resolving.de/urn:nbn:de:bvb:19-109650>
- **Silling, Sebastian** (2011): Einführung in das Google Web Toolkit am Beispiel eines Projektes aus der Verkehrstelematik. Online verfügbar unter: URL: [https://www.matse.itc.rwth-aachen.de/dienste/public/show\\_document.php?id=8098](https://www.matse.itc.rwth-aachen.de/dienste/public/show_document.php?id=8098)

# List of figures

- Slide 3, <http://www.wittgensteinsource.org/>
- Slide 9, created with Unitex application
- Slide 12, [http://maxdemo.cis.uni-muenchen.de/home\\_demos/cisweb/Manual.pdf](http://maxdemo.cis.uni-muenchen.de/home_demos/cisweb/Manual.pdf)
- Slide 14, <http://www.mediaclick.de/wp-content/uploads/2014/01/java.png> , <http://everflux.de/wp-content/uploads/2008/04/gwt.png>  
<http://polpix.sueddeutsche.com/bild/1.1072372.1355681892/860x860/debatte-urheberrecht.jpg>
- Slide 19, <http://www.mediabistro.com/alltwitter/files/2013/12/message-app.jpg>

# WiTTFind

<http://wittfind12.cis.uni-muenchen.de>

**CIS**      CENTRUM FÜR INFORMATIONS UND SPRACHVERARBEITUNG  
WITTGENSTEIN ARCHIVES UNIVERSITY OF BERGEN      **WAB**

Regelbasiertes Finden    Semantisches Finden    Ohne Alternativen Finden    Graphischer Editor    Geheimschriftübersetzer

Hilfe

WiTTFind (Developer-Version)

WiTTFind-Suche

Regelbasiertes, linguistisches Finden auf dem Big Typescript TS-213 von L. Wittgenstein

Für nähere Informationen zum Projekt:

[hier klicken.](#)

Webseite: © Wittfind Version 1.0.0

WAB - The Wittgenstein Archives at the University of Bergen & CIS - Centrum für Informations- und Sprachverarbeitung.