

Integration des Graphischen Editors CISGraph für Lokale Grammatiken in die Suchmaschine WiTTFind

Centrum für Informations- und Sprachverarbeitung
LMU München

Yuliya Kalasouskaya
Betreuer: Dr. Maximilian Hadersbeck

02.06.2014

Gliederung

1. Einführung
2. Lokale Grammatiken als Grundlage für CISGraph
3. Unitex und CISGraph
4. CISGraph in CISWeb 2.0
5. Google Web Toolkit
6. Client-Server Kommunikation
7. HTML5 Storage Konzept
8. Graphische Oberfläche
9. Testen und Dokumentation
10. Zusammenfassung
11. Ausblick
12. Literaturverzeichnis

Einführung

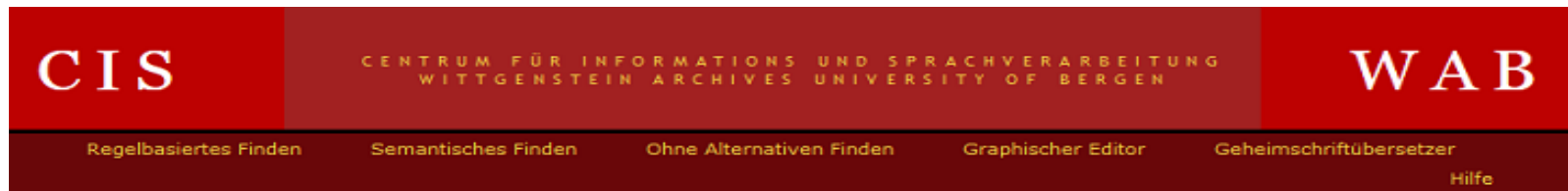
- Wittgenstein Scholarship and Wittgenstein in Co-Text am CIS
- WiTTFind ist eine Suchmaschine, die für die Suche in Texten von Ludwig Wittgenstein eingesetzt wird



Einführung

Hauptziele der Bachelorarbeit:

- Erweiterung von WiTTFind um den graphischen Editor CISGraph
- Entwicklung von lokalen Grammatiken, die als Suchanfrage für WiTTFind verwendet werden



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

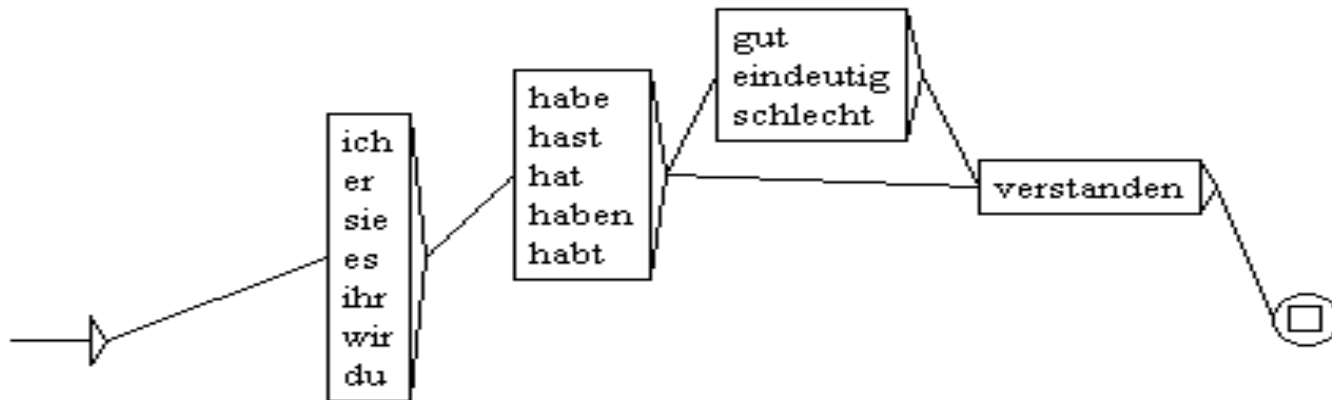
WAB - The Wittgenstein Archives at the University of Bergen & CIS - Centrum für Informations- und Sprachverarbeitung.

Gliederung

1. Einführung
2. Lokale Grammatiken als Grundlage für CISGraph
3. Unitex und CISGraph
4. CISGraph in CISWeb 2.0
5. Google Web Toolkit
6. Client-Server Kommunikation
7. HTML5 Storage Konzept
8. Graphische Oberfläche
9. Testen und Dokumentation
10. Zusammenfassung
11. Ausblick
12. Literaturverzeichnis

Lokale Grammatiken als Grundlage für CISGraph

- Sprache als Menge aller Sätze, deren Struktur und Aufbau bestimmten Regeln unterliegen
- Repräsentation durch reguläre Ausdrücke, Ersetzungsregeln oder als Graph:



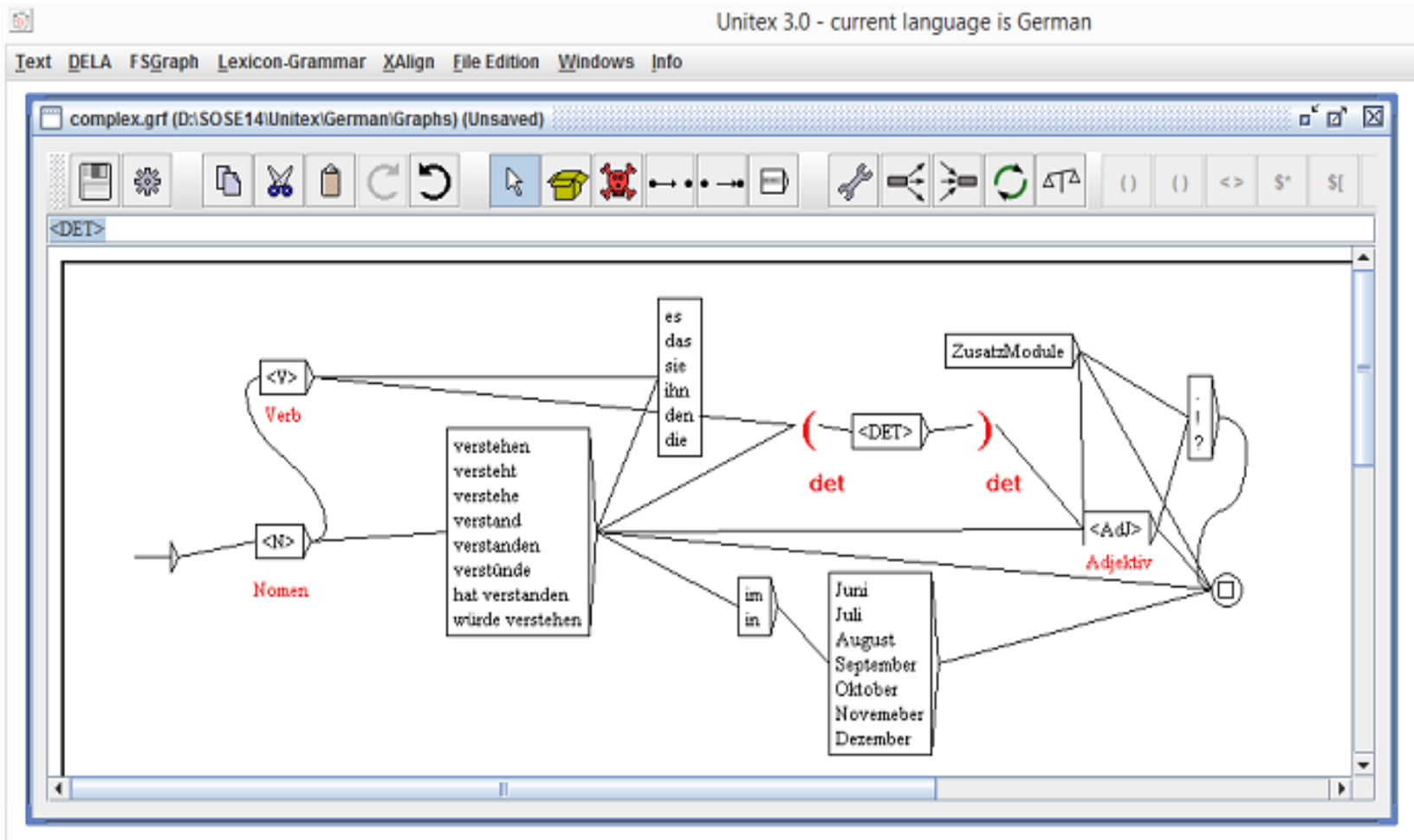
Gliederung

1. Einführung
2. Lokale Grammatiken als Grundlage für CISGraph
- 3. Unitex und CISGraph**
4. CISGraph in CISWeb 2.0
5. Google Web Toolkit
6. Client-Server Kommunikation
7. HTML5 Storage Konzept
8. Graphische Oberfläche
9. Testen und Dokumentation
10. Zusammenfassung
11. Ausblick
12. Literaturverzeichnis

Unitex

- Open Source Korpustool
- Entwicklung am Gaspard-Monge Institut der Université de Marne la Vallée
- Einsatzbereiche:
 - Verwendung elektronischer Wörterbücher
 - Erstellung lokaler Grammatiken
 - Verwendung regulärer Ausdrücke bei der Suche

Unitex



Unitex und CISGraph

Unitex



Ausführbare Java-Anwendung



Betriebssystemunabhängig



Lokale Speicherung auf der Festplatte

CISGraph



Webbasierte Anwendung



Betriebssystemunabhängig



(für diese Bachelorarbeit relevant)
Lokale Speicherung im Browser und auf
der Festplatte

Gliederung

1. Einführung
2. Lokale Grammatiken als Grundlage für CISGraph
3. Unitex und CISGraph
- 4. CISGraph in CISWeb 2.0**
5. Google Web Toolkit
6. Client-Server Kommunikation
7. HTML5 Storage Konzept
8. Graphische Oberfläche
9. Testen und Dokumentation
10. Zusammenfassung
11. Ausblick
12. Literaturverzeichnis

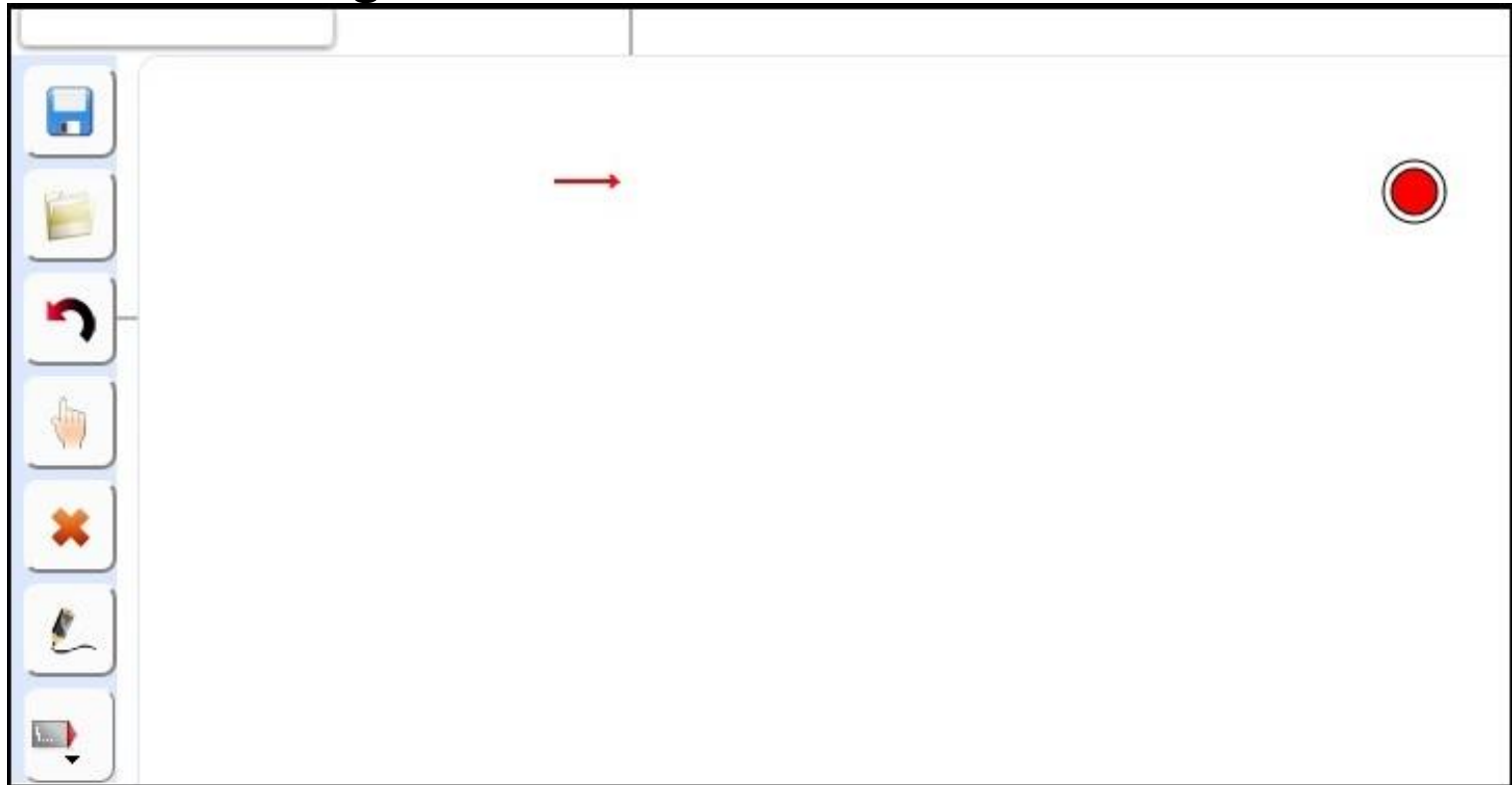
CISGraph in CISWeb 2.0

CISWeb 2.0 Toolsammlung:



CISGraph in CISWeb 2.0

- Entwicklung 2012 von Shuangjiao Cao
- Unterstützung durch Patrick Seebauer



Gliederung

1. Einführung
2. Lokale Grammatiken als Grundlage für CISGraph
3. Unitex und CISGraph
4. CISGraph in CISWeb 2.0
- 5. Google Web Toolkit**
6. Client-Server Kommunikation
7. HTML5 Storage Konzept
8. Graphische Oberfläche
9. Testen und Dokumentation
10. Zusammenfassung
11. Ausblick
12. Literaturverzeichnis

Google Web Toolkit (GWT)

- 2006 Veröffentlichung von Google
- Entwicklung und Optimierung von Webanwendungen
- Java zu JavaScript Umwandlung



Verwendete Quellen:

<http://www.mediatick.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>

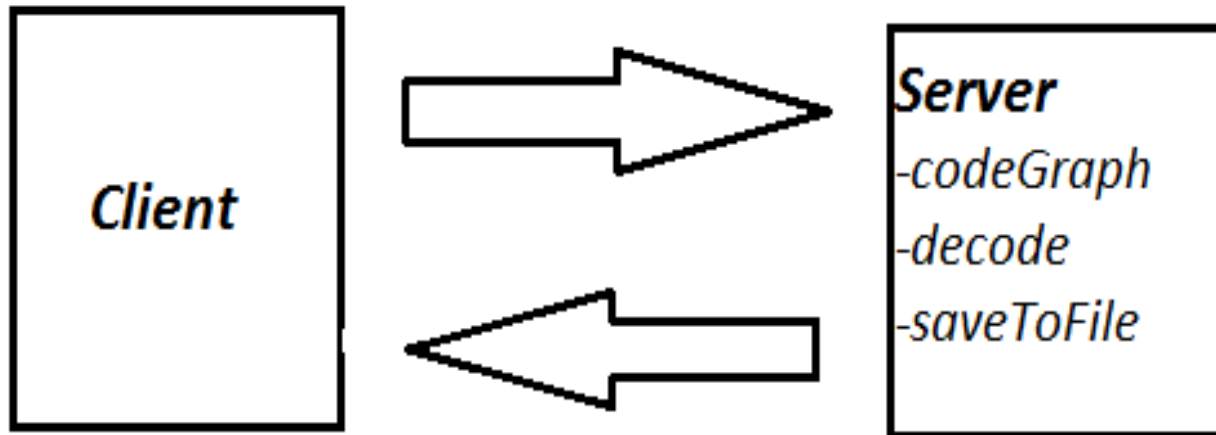
Google Web Toolkit (GWT)

- Graph_editor 687 [svn://svn.cis.uni-muenchen.de/cisv
 - src 687
 - com.hydro4ge.raphaelgwt 237
 - com.hydro4ge.raphaelgwt.client 237
 - de.cis.grapheditor 687
 - de.cis.grapheditor.client 687
 - de.cis.grapheditor.client.external_lib 685
 - de.cis.grapheditor.server 237
 - de.cis.grapheditor.shared 667
 - test 3
 - JRE System Library [JavaSE-1.7]
 - GWT SDK [GWT - 2.6.0]
 - Ivy ivy.xml [*]
 - JUnit 4
 - Referenced Libraries
 - bin 652
 - doc -1
 - gwt-unitCache
 - lib
 - war 672
 - www-test 3
 - build.xml 203
 - ivy.xml 650



Google Web Toolkit (GWT)

Ursprüngliche Implementierung:



Google Web Toolkit (GWT)

Änderungen und neue Entwicklung:

- Implementierung des GWT-Projektes ohne eingebauten GWT-Server
- Datenverwaltung auf der Clientseite
- Implementierung von neuen Komponenten
- CISGraph Integration in andere Projekte

Gliederung

1. Einführung
2. Lokale Grammatiken als Grundlage für CISGraph
3. Unitex und CISGraph
4. CISGraph in CISWeb 2.0
5. Google Web Toolkit
- 6. Client-Server Kommunikation**
7. HTML5 Storage Konzept
8. Graphische Oberfläche
9. Testen und Dokumentation
10. Zusammenfassung
11. Ausblick
12. Literaturverzeichnis

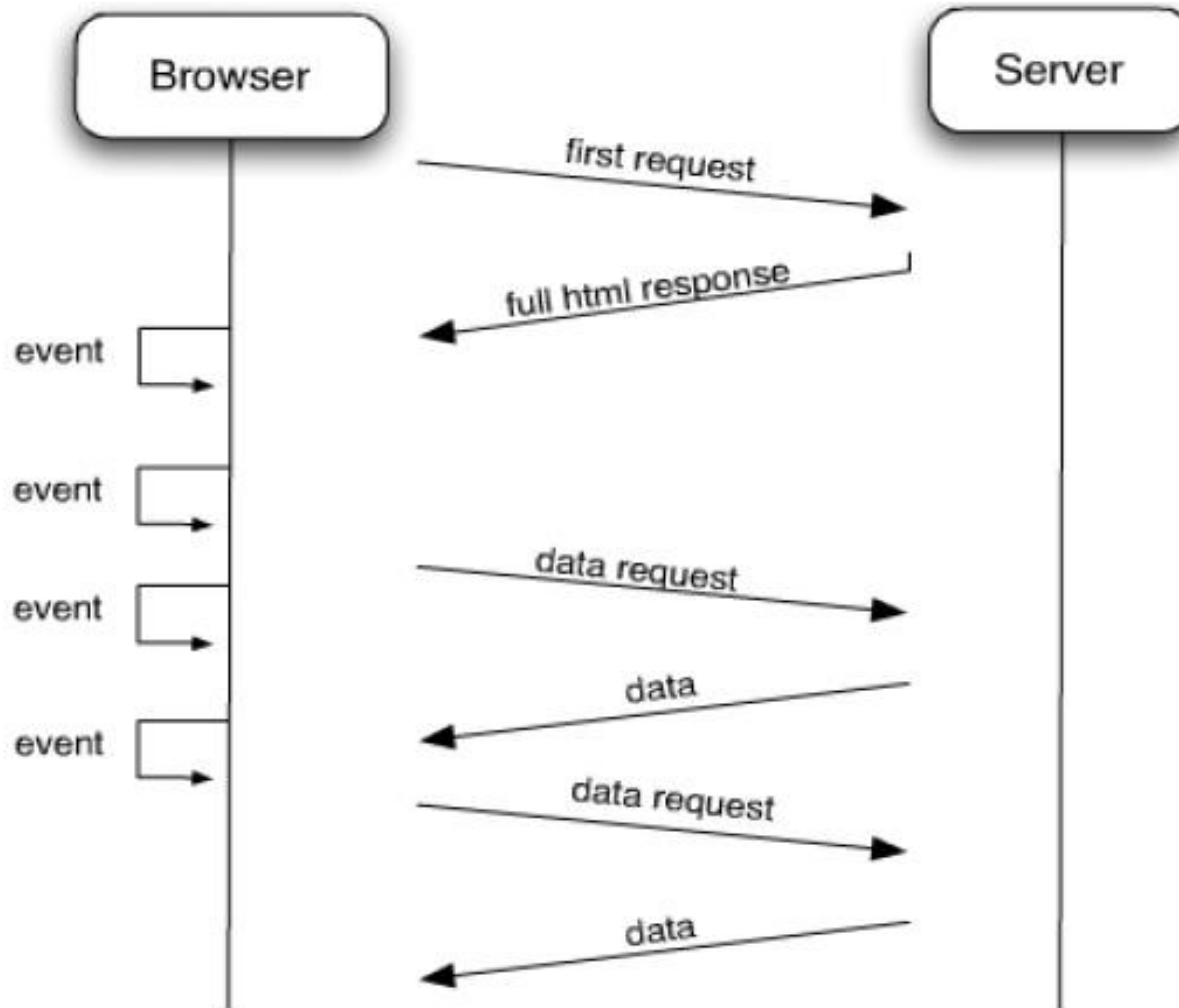
Client-Server Kommunikation

- Kommunikation anhand von JSON Nachrichten
- JSON Objekt ist eine Zeichenkette, die die Beschreibung eines Graphen enthält:

```
"{"Nodes":{"0":{"type":"START", "x":200, "y":60, "data":{"subgraph":null}}, "1":{"type":"FINAL", "x":600, "y":50, "data":{"subgraph":null}}, "2":{"type":"MATCH", "x":423, "y":151, "data":"lesen"}}, "Links":[{"from":0, "to":2}, {"from":2, "to":1}]}"
```

- Die Umwandlung des Graphen in ein JSON Objekt und die Dekodierung des JSON Objektes zum Graphen erfolgt auf der Clientseite
- METHOD_POST ermöglicht es, den kompletten Graphen an die angegebene URL- Adresse abzuschicken.

Client-Server Kommunikation



Client-Server Kommunikation

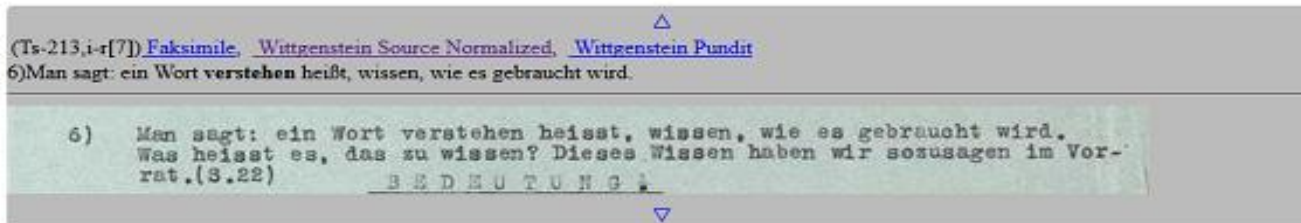
WiTTFind schickt eine HTML-Seite als Antwort:



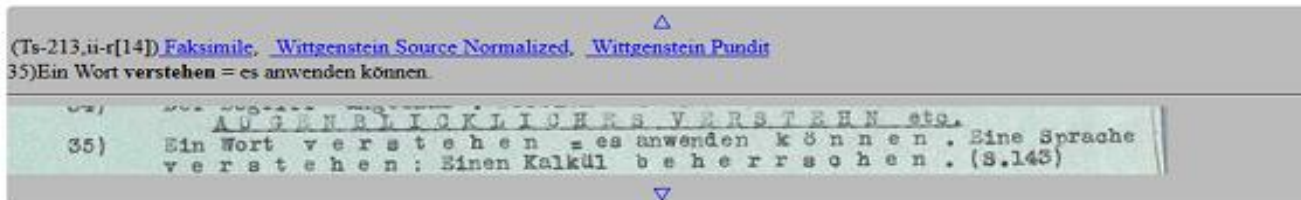
The header of the website features a red background with three main sections. On the left, the letters 'CIS' are displayed in white. In the center, the text 'CENTRUM FÜR INFORMATIONEN UND SPRACHVERARBEITUNG WITTGENSTEIN ARCHIVES UNIVERSITY OF BERGEN' is written in small white capital letters. On the right, the letters 'WAB' are displayed in white. Below this header, a dark red navigation bar contains four links: 'Regelbasiertes Finden', 'Semantisches Finden', 'Ohne Alternativen Finden', and 'Hilfe'.

WiTTFind WiTTFind-Suche

318 Treffer gefunden



This snippet shows a search result for the word 'verstehen'. It includes a small upward-pointing triangle icon, a list of links: '(Ts-213,i-r[7]) [Faksimile](#), [Wittgenstein Source Normalized](#), [Wittgenstein Pundit](#)', and the text '6)Man sagt: ein Wort **verstehen** heißt, wissen, wie es gebraucht wird.' Below this is a light green box containing a scanned document snippet: '6) Man sagt: ein Wort verstehen heisst, wissen, wie es gebraucht wird. Was heisst es, das zu wissen? Dieses Wissen haben wir sozusagen im Vorrat.(S.22) BEDEUTUNG:'.



This snippet shows a search result for the word 'verstehen'. It includes a small upward-pointing triangle icon, a list of links: '(Ts-213,ii-r[14]) [Faksimile](#), [Wittgenstein Source Normalized](#), [Wittgenstein Pundit](#)', and the text '35)Ein Wort **verstehen** = es anwenden können.' Below this is a light green box containing a scanned document snippet: '35) AUGENBLICKLICHERS VERSTEHEN etc. Ein Wort verstehen = es anwenden können. Eine Sprache verstehen: Sinen Kalkül beherrschen. (S.143)'.

Gliederung

1. Einführung
2. Lokale Grammatiken als Grundlage für CISGraph
3. Unitex und CISGraph
4. CISGraph in CISWeb 2.0
5. Google Web Toolkit
6. Client-Server Kommunikation
- 7. HTML5 Storage Konzept**
8. Graphische Oberfläche
9. Testen und Dokumentation
10. Zusammenfassung
11. Ausblick
12. Literaturverzeichnis

HTML5 Storage Konzept

- Speicherung von Daten im Browser
- Typen von Storage:
 - localStorage
 - sessionStorage
- Vorteile localStorage:
 - Keine Belastung des Netzverkehrs
 - Kein Einfluss auf die Wiedergabe bei Verzögerungen auf der Serverseite
 - Freier Zugriff auf die vorher geladenen Dateien im Browser
 - Problemlose Speicherung von Daten auf der Festplatte
 - Beibehalten von Daten nach Browser-Neustart

Gliederung

1. Einführung
2. Lokale Grammatiken als Grundlage für CISGraph
3. Unitex und CISGraph
4. CISGraph in CISWeb 2.0
5. Google Web Toolkit
6. Client-Server Kommunikation
7. HTML5 Storage Konzept
- 8. Graphische Oberfläche**
9. Testen und Dokumentation
10. Zusammenfassung
11. Ausblick
12. Literaturverzeichnis

Graphische Oberfläche

CIS

CENTRUM FÜR INFORMATIONEN UND SPRACHVERARBEITUNG
WITTGENSTEIN ARCHIVES UNIVERSITY OF BERGEN

WAB

Regelbasiert Finden Graphischer Editor Semantische Suche Ohne Alternativen Suche Wittgenstein Source Hilfe

<-neu hinzugefügt

**<-/
abgeändert**

**<-/
neu hinzugefügt**

**<-/
neu hinzugefügt**

**<-/
neu hinzugefügt**

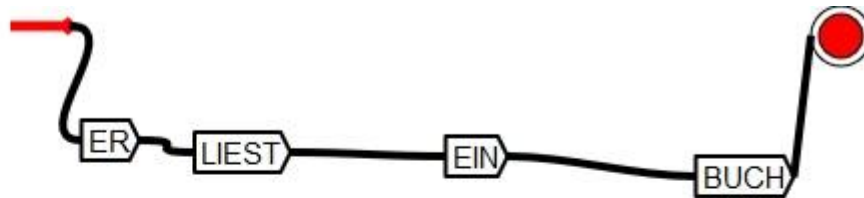
HIER IST EIN TOLLER GRAPH

< >

You can select and move boxes by clicking on them.

Graphische Oberfläche

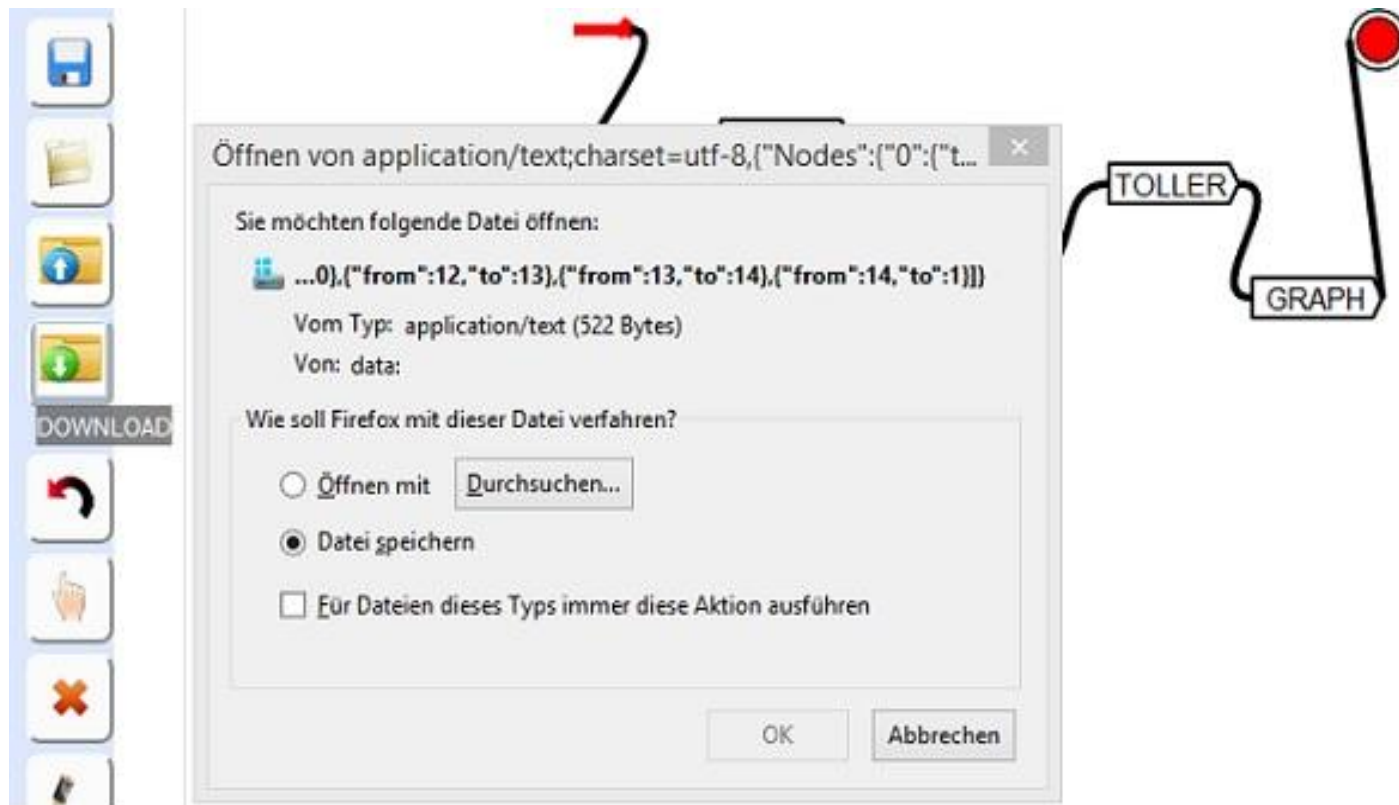
Speicherung im Browser:



Einen Graphnamen eingeben!

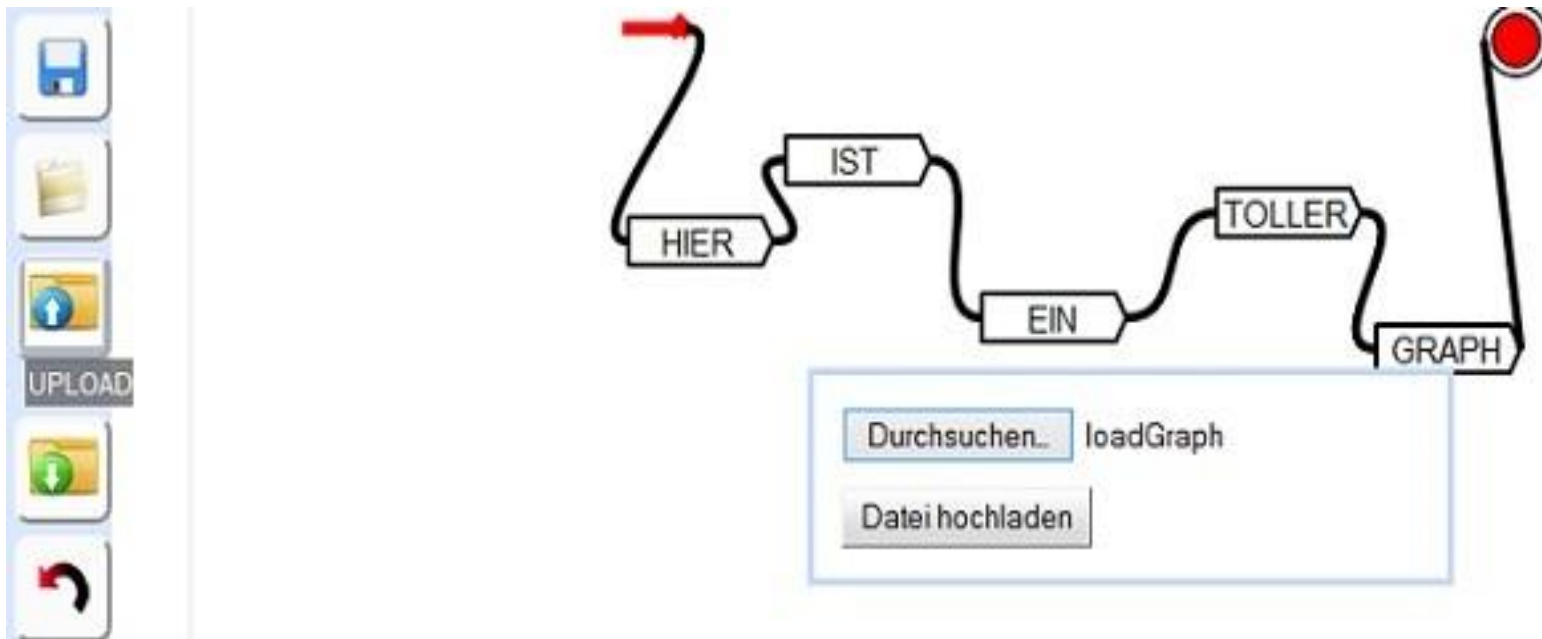
Graphische Oberfläche

Speicherung auf der Festplatte:



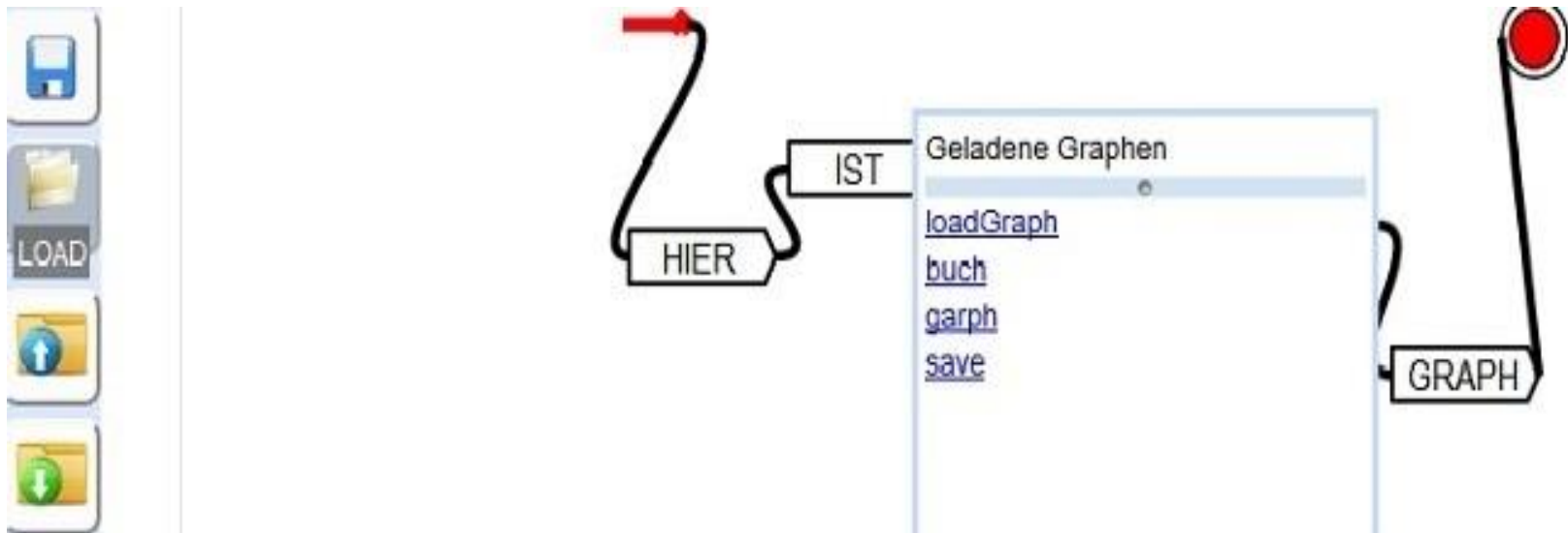
Graphische Oberfläche

Laden von der Festplatte in den Browser:







Graphische Oberfläche

Laden vom Browser in die graphische Oberfläche:



Graphische Oberfläche

Allgemeine Vorgehensweise:

-  Ein Graph wird erstellt und im Browser gespeichert
-  Mit DOWNLOAD Funktion wird der erstellte Graph auf den Rechner heruntergeladen
-  Mit UPLOAD kann ein bestehender Graph vom Rechner in den Browser geladen werden
-  Für die Darstellung des Graphen auf der graphischen Oberfläche, muss der LOAD Knopf benutzt werden.

Gliederung

1. Einführung
2. Lokale Grammatiken als Grundlage für CISGraph
3. Unitex und CISGraph
4. CISGraph in CISWeb 2.0
5. Google Web Toolkit
6. Client-Server Kommunikation
7. HTML5 Storage Konzept
8. Graphische Oberfläche
- 9. Testen und Dokumentation**
10. Zusammenfassung
11. Ausblick
12. Literaturverzeichnis

Testen

- Komponententest
- Einsatz von JUnit Tests
- Name der zu testenden Klasse + Test

```
@Test
public void testCreateNewKeyForGraph() {
    kp.createNewKeyForGraph("myKey");
    Assert.assertFalse(kp.getSetOfKeys().isEmpty());
}
```

```
@Test(expected=IllegalArgumentException.class)
public void testSetTypeNullIsNotPossibleValue() {
    node.setType(null);
}
```

Dokumentation

- Ausführliche und einheitliche Dokumentation
- Einsatz von Javadoc
- Beispiel der Methodenbeschreibung:

```
/**  
 * Gets the actual storage object.  
 *     if local or session storage are supported->initialize the object  
 *     else throw Exception  
 * @param type of storage object  
 * @return storage object  
 * @throws Exception if the given type of storage is not supported  
 * @see de.cis.grapheditor.client.IJSONReader  
 * @see com.google.gwt.storage.client.Storage  
 */  
public Storage getStorage(Type type) throws Exception{
```

Dokumentation

All Classes

Packages

com.hydro4ge.rafaelgwclient
de.cis.grapheditor.client
de.cis.grapheditor.client.external_lib
de.cis.grapheditor.shared

All Classes

AbortEvent
AbortHandler
AnimationCallback
Attr
Attr.StrokeDashArray
Attr.StrokeLineCap
Attr.StrokeLineJoin
Attr.TextAnchor
BBox
Blob
BoxDoubleClicked
BoxDoubleClickedEvent

Overview Package Class Use Tree Deprecated Index Help

Prev Next Frames No Frames

Packages

Package	Description
---------	-------------

com.hydro4ge.rafaelgwclient	
-----------------------------	--

de.cis.grapheditor.client	
---------------------------	--

de.cis.grapheditor.client.external_lib	
--	--

de.cis.grapheditor.shared	
---------------------------	--

Overview Package Class Use Tree Deprecated Index Help

Prev Next Frames No Frames

Gliederung

1. Einführung
2. Lokale Grammatiken als Grundlage für CISGraph
3. Unitex und CISGraph
4. CISGraph in CISWeb 2.0
5. Google Web Toolkit
6. Client-Server Kommunikation
7. HTML5 Storage Konzept
8. Graphische Oberfläche
9. Testen und Dokumentation
- 10. Zusammenfassung**
11. Ausblick
12. Literaturverzeichnis

Zusammenfassung

- Entstehung der CISGraph- Anwendung und Integration dieser Anwendung in die Suchmaschine WiTTFind
- GWT Projekt ohne GWT Server
- Datenverarbeitung und Datenspeicherung auf der Clientseite
- HTML5 Storage Einsatz
- JUnit Tests und Dokumentation

Ausblick

- *Komplexere Anfragen*
- *Namensänderung der Graphendatei während der Speicherung auf Festplatte*
- *Aktueller Graph als Screenshot im Trefferfenster*
- *Unterordner für Subgraphen im LocalStorage*
- *Datenbankenstruktur für vordefinierte Graphen*

Literaturverzeichnis

- **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

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.