



"Digital Humanities im deutschsprachigen Raum"
die zweite Jahrestagung vom 23.-27.2.2015 in Graz

**Wittgensteins Nachlass: Erkenntnisse und Weiterentwicklung der
FinderApp WiTTFind**

Max Hadersbeck, Alois Pichler, Florian Fink, Daniel Bruder,
Ina Arends, Johannes Baiter

Centrum für Informations- und Sprachverarbeitung (CIS), LMU, München,
Wittgenstein Archiv, Universität Bergen (WAB), Norwegen.

Die Entwicklung der FinderApp WiTTFind

Ludwig Wittgenstein (1889-1951) hinterließ 20,000 Seiten: Wittgenstein's Nachlass

2000: Wittgenstein Archives, University Bergen (WAB):

Nachlass CD-ROM Edition, Bergen Electronic Edition (BEE), Windows NT

2009: WAB und net7 (Pisa)

-> Wittgenstein Source, 5000 Seiten, Standard Such-Tools

2010: WAB und CIS: "Wittgenstein in Co-Text":

Philosophie und Computerlinguistik

2011: WEB-basierte FinderApp WiTTFind, erste Version

Vorstellung auf Wittgenstein Symposium, Kirchberg

2014: **WiTTFind Gewinner des EU-AWARD**

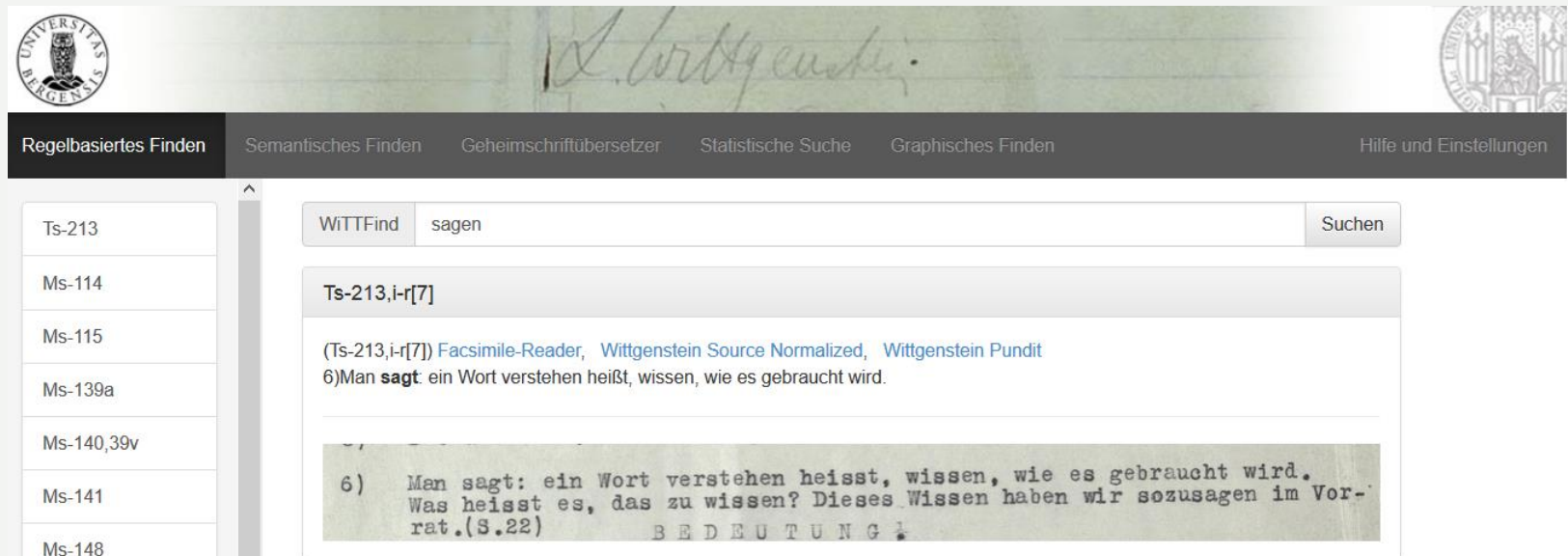
EU-Projekt Digitised Manuscripts to Europeana (DM2E),

Vorstellung in Pisa, 2014

FinderApp WiTTFind

Regelbasierte WEB-Applikation zum Finden von Textstellen im Nachlass von
Ludwig Wittgenstein

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



UNIVERSITAS
MUNICHENSIS

Regelbasiertes Finden Semantisches Finden Geheimschriftübersetzer Statistische Suche Graphisches Finden Hilfe und Einstellungen

Ts-213
Ms-114
Ms-115
Ms-139a
Ms-140,39v
Ms-141
Ms-148

WITTFind sagen Suchen

Ts-213,i-r[7]

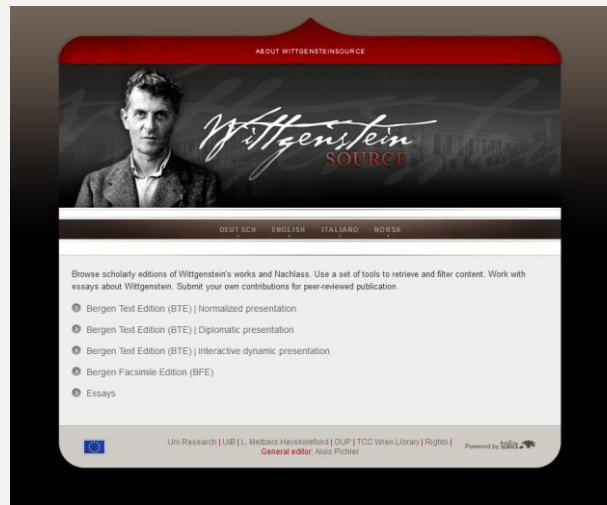
(Ts-213,i-r[7]) Facsimile-Reader, Wittgenstein Source Normalized, Wittgenstein Pundit

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!

Grundlagen von WiTTFind: Texte

5000 Seiten, öffentlich zugänglicher Teil des Nachlasses (1951)
vom Wittgenstein Archiv in Bergen/Norwegen, Bergen Electronic Edition
TEI-P5 konformes XML-Format, Dr. A. Pichler, Ø. Gjesdal
<http://www.wittgensteinsource.org/>



Passau 2014: John Nerbonne: “Langsames Vorgehen bei großen Fragen”

Grundlagen von WiTTFind: Lexikon

- elektronisches, multilinguales Vollformen- und Phrasenlexikon (WiTTLex)
morphologische, syntaktische und semantische Kategorien, CISLEX
53395 Einträge
- Part of Speech Tagger
treetagger, Dr. Schmid (CIS)
- regelbasierter Zugriff,
- lokale Grammatiken,
- satzsegmentierte Analyse

```

1 nachdenken, nachdenken.V+#4
2 farbgleiche, farbgleich.ADJ+COL+KOMP
3 part of the practice of communication, .X+z1
4 part of the time, .X+z1
5 nachdenkender, nachdenkend.ADJ+ER
6 rot, rot.ADJ+COL+Grundfarbe
7 Russell, ..EN:geM
8 Russells, Russell.EN:geM
9 Russellsche, Russell.EN
10 Russellschen, Russell.EN
11 Russen, Russe.N:aeM:amM:deM:dmM:geMS:gmM:nmM

```

Wittgenstein: Tractatus logico philosophicus [22, 3.3]:

„Nur der Satz hat Sinn; nur im Zusammenhang des Satzes hat ein Name Bedeutung“.

Grundlagen von WiTTFind: Faksimile bei Trefferanzeige

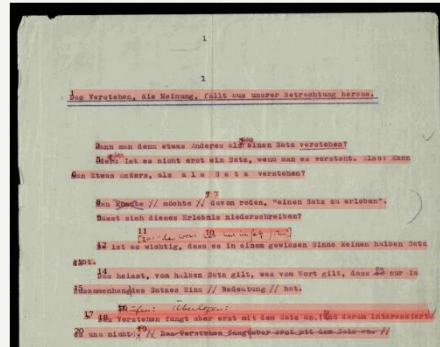
Multiuserfähiges, halbautomatisches WEB-basiertes OCR Korrektur Tool

(R.Capsamun)

OCR mit tesseract

Typoskripte

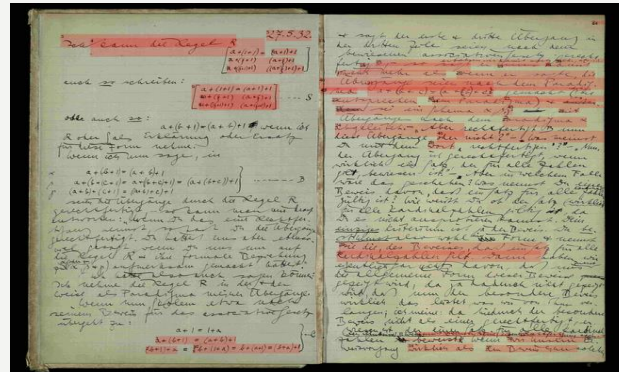
(recht gut)



- 1: Das Verstehen die Meinung fällt aus unsrer Betrachtung häraus
- 2: iwü
- 3: Kann man dann etwas Änderes ulsTeinen Satz verstgggg
- 4: 1 634
- 5: Odeñh ist es nicht erst ein Satz wenn man es versteht 150 Kann
- 6: män Etwas gnders, als a l s 3 a t z verstehen
- 7: S 7
- 8: Man gynä3e möchte f davon reden einen Satz zu erleben
- 9: Lasst sich dieses ärlebnis niederschreiben
- 10: , E h
- 11: fpxti Lnnv LAuuh JLÄ fhj
- 12: Da ist es wichtig dass es in einem gewissen Sinne Keinen halben Satz
- 13: gibt
- 14: Das heisst vom hxlben Satz gilt was vom hort gilt äassfü»nur im
- 15: zusammenhanääes Satzes Sinn Bedeutung hat
- 16: Qhagw am i Ä

Manuskripte

(sehr schlecht)



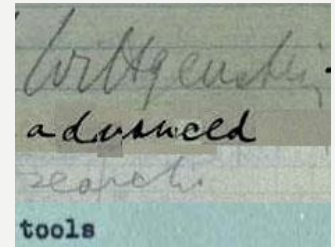
- 1; a4 CH aFfvÖf
- 2; ßHfic»
- 3; w w
- 4; cuMj
- 5; z 7 552
- 6; auVI
- 7; a f I aaf ,L
- 8; a 43700 fdfffljH
- 9; 4 aJrljIiY
- 10; cL a p

Grundlagen von WiTTFind: Opensource Software

Wittgenstein Advanced Search Tools (WAST)

(F. Fink, D. Bruder, S. Schweter, M. Lindinger, Y. Kalasousayka)

- Versionierung mit gitlab,
siehe: <https://gitlab.cis.uni-muenchen.de>
- Test driven Development (TDD),
- Continuous Integration (CI),
- Feedback-app für Rückmeldung und Wartung
- Virtualisierte Programme: Linux, Mac, Windows
- docker-Technologie (see www.docker.com).
- WEB-Programme Browser unabhängig: bootstrap
- Rich-clients HTML5, local storage



Feedback

Email

email... email required

Component

Please choose component component required

Short description

Short description... Short description needed

Severity

Severity slider

Important Issue

Description

What happened? How to reproduce? ...

Please login to submit a report!

Die Oberfläche von WiTTFind

<http://wittfind.cis.uni-muenchen.de> (S. Schweter)

The screenshot shows the WiTTFind web interface. At the top, there is a navigation bar with the following options: Regelbasiertes Finden, Semantisches Finden, Geheimschriftübersetzer, Statistische Suche, Graphisches Finden, and Hilfe und Einstellungen. On the left side, there is a vertical list of manuscript identifiers: Ts-213, Ms-114, Ms-115, Ms-139a, Ms-140,39v, Ms-141, Ms-148, Ms-149, Ms-150, Ms-152, and Ms-153. The main content area displays a search result for 'Man könnte sagen'. The result includes the identifier 'Ts-213,1r[7]' and links to 'Facsimile-Reader', 'Wittgenstein Source Normalized', and 'Wittgenstein Pundit'. The text of the result is: 'Man könnte sagen: Was soll uns das Verstehen kümmern? Wir müssen ja den Satz verstehen, daß er für uns ein Satz ist!'. Below the text is a facsimile image of the original manuscript page, which shows the handwritten text 'Man könnte sagen: Wir haben es also in unsern Betrachtungen mit dem Verstehen des Satzes nicht zu tun; denn wir selbst müssen ihn verstehen, damit er für uns ein Satz ist. Was soll uns da Verstehen kümmern? Wir müssen ja den Satz verstehen, daß er für uns ein Satz ist!' with a checkmark next to the first sentence.

Spezielle Features von WiTTFind: Lemmatisierte Vorschlagssuche (J. Baiter)

Eintippen von: ..stehen...

verstehen ->

verstehe ->

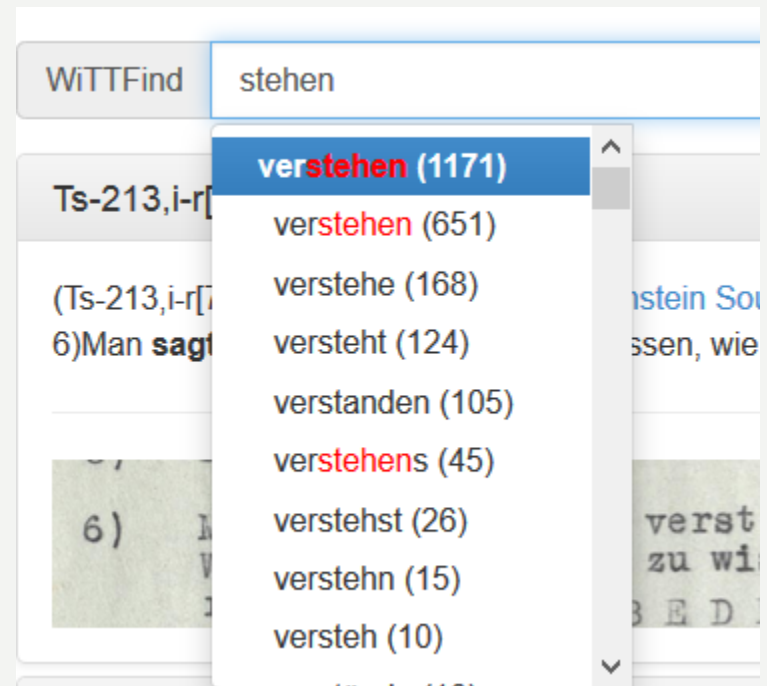
versteht ->

versteh usw.

Anzeige mit Frequenz im Nachlass

Basierend auf symmetrischem
Index und Autovervollständigung.

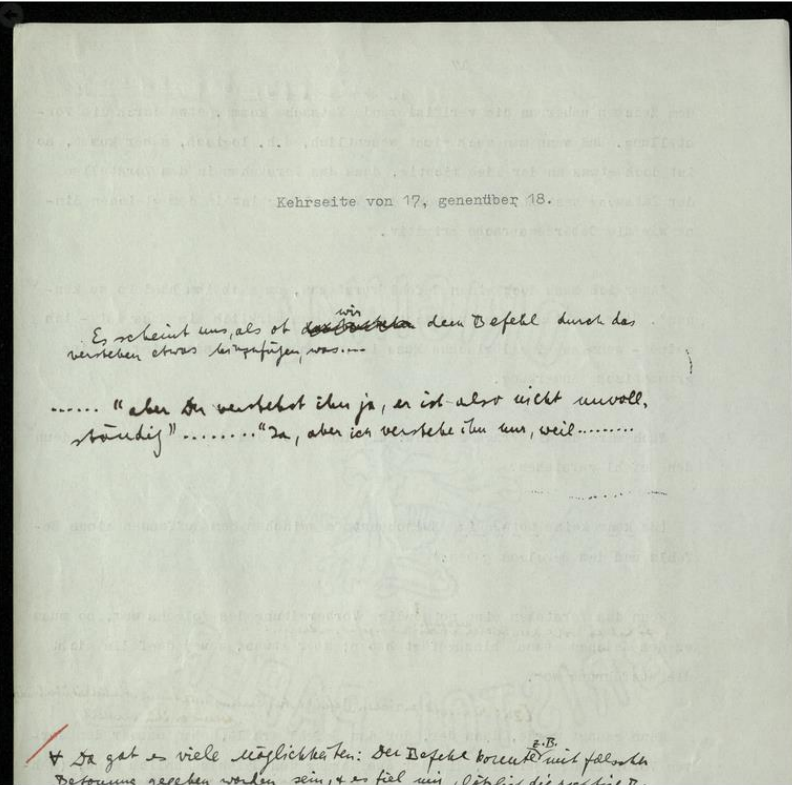
SIS von Daniel Bruder, (see: <http://sis.cis.lmu.de>)



Spezielle Features von WiTTFind:

Faksimile-Reader mit Treffer-Highlighting und blättern (M. Lindinger)

Vorheriger Treffer Nächster Treffer Vollbild öffnen HD good quality Hilfe



Erscheinung, als ob ^{ein} ~~der~~ Befehl durch das verstehen etwas hinzugefügt, was...

..... "aber Du verstehst ihn ja, er ist also nicht unvollständig" "Ja, aber ich verstehe ihn nur, weil....."

Es gibt so viele Möglichkeiten: Der Befehl könnte mit jeder Bedeutung gedeutet werden sein, es fällt mir nicht ein, die welche...

18

grössere Ähnlichkeit ersetzt. Und der Weg vom Symbol zur Wirklichkeit scheint hier // nun // sehr verkürzt zu sein. (Schemata könnte ich, um zu beschreiben, in welcher Stellung ich mich bei der und der Gelegenheit befinden habe, diese Stellung einnehmen.) (Richt: Erwarten, Wünschen, etc.)

Es ist damit auch gezeigt, dass ^{(dieses Phantasiebildes) Vorstellungen} das Vorkommen von Phantasiebildern, Vorstellungen, für den Gedanken ~~das~~ unwesentlich ist. // ^{nicht} Es ist damit auch das Unwesentliche der Phantasiebilder für den Gedanken gezeigt //

[siehe § 89/4] Zu: "Sehen" ?

Ich könnte auch sagen: Es scheint uns, als ob, wenn wir den Befehl * (z.B. $\frac{x}{x^2} 1 2 3$) # verstehen, wir etwas hinzufügen, was die Lücke füllt. ^{zur vollen Befehls-Ausführung} ^{was ich ja vollkomme}

So dass wir dem, der sagt "aber Du verstehst ihn ja" antworten können: "Ja, aber nur, weil ich noch etwas hinzufüge: die Deutung nämlich." (nicht: Erwartung etc.)

Zu: "Sehen"

Nun müsste man allerdings darauf sagen: Aber was veranlasst Dich denn zu gerade d i e s e r Deutung? Ist es der Befehl, dann war er ja schon eindeutig, da er nur diese Deutung befahl. Oder, hast Du die Deutung willkürlich hinzugefügt - , dann hast Du ja auch den Befehl nicht verstanden, sondern erst das, was Du aus ihm [auf eigene Faust] gemacht hast. ^{Sieh gerade zu diesem}

Eine 'Interpretation' ist doch wohl etwas, was in Worten gegeben wird. ^{Zeichen}

Es ist d i e s e Interpretation im Gegensatz zu einer anderen (die anders lautet). - Wenn man also sagt "jeder Satz bedarf noch einer Interpretation", so hiesse das: kein Satz kann ohne einen Zusatz verstanden werden. ^{was ich will}

Spezielle Features von WiTTFind:

Ausführliche Hilfeseiten mit Beispielen und Videotutorial (A. Krey)

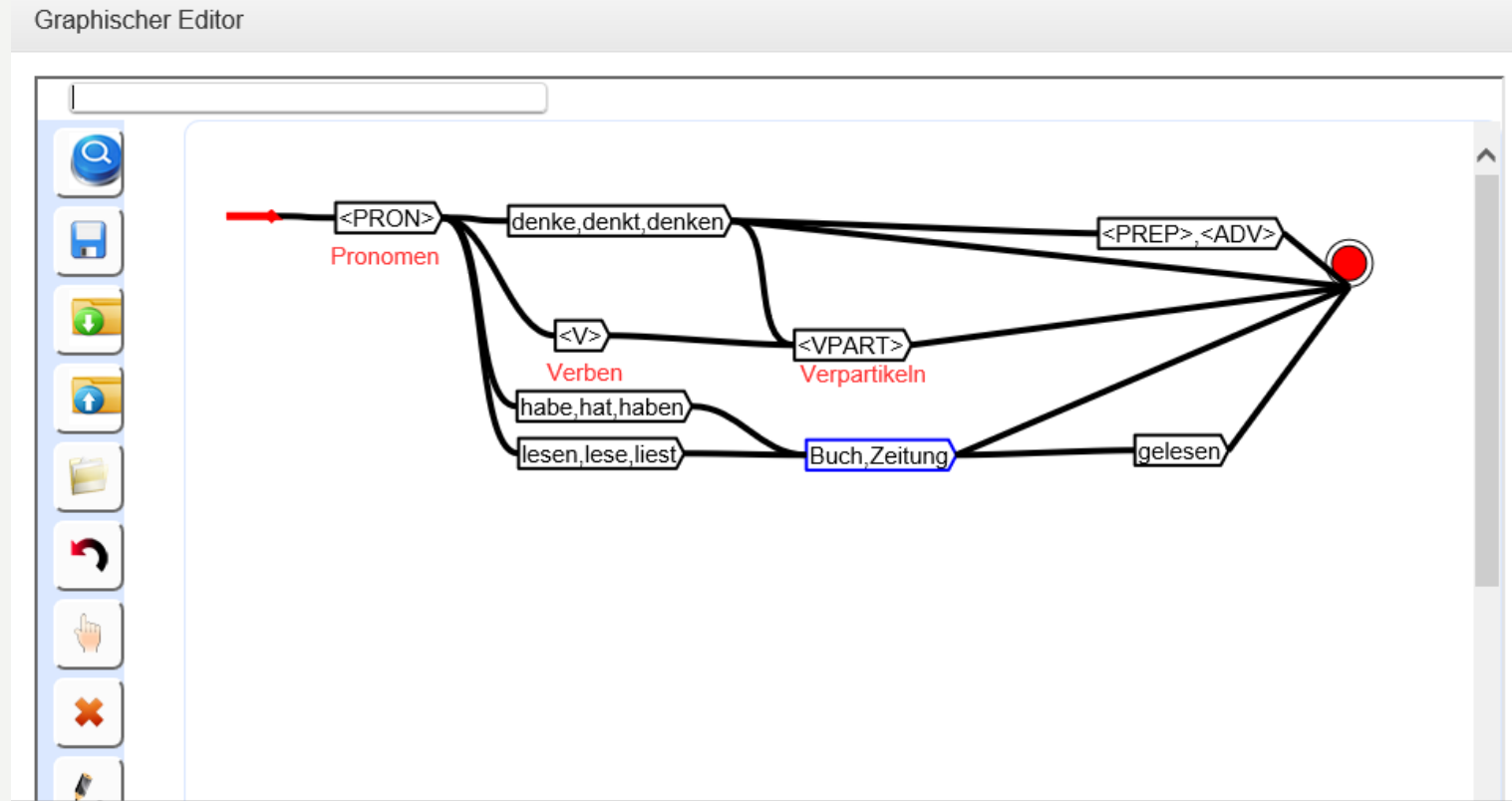
The screenshot shows the WiTTFind web interface. At the top, there is a navigation bar with links: [Regelbasiertes Finden](#), [Semantisches Finden](#), [Geheimschriftübersetzer](#), [Statistische Suche](#), [Graphisches Finden](#), and [Hilfe und Einstellungen](#). The main content area is titled "WITTFind" and shows a search for the word "stehen".

The search results are organized into several sections:

- einfache Suche nach Wörtern**: Details show the word "schwarz" with its morphological variants: "schwarzer", "schwarz Kreis", "schwarzen" Kreis, "der Kreis", "der * Kreis", and "der Kr*".
- Satzkategorien**: Details show the sentence "mir <PUNCT> was mir <PUNCT> was <PUNCT> <MOT> <PUNCT> <BOS> ich können <PUNCT> <EOS>" with its morphological variants: "<BOS> <MOT> <EOS>".
- Lexikalische Wortkategorien**: Details show the sentence "<EN> und <EN> <KONJ> <EN> die <ADJ> Farbe die gleiche Farbe die gleiche <N> die gleiche <N> <V> <MOT> <EN>" with its morphological variants.
- Lexikalische Wortkategorien um morphologische verfeinert**: Details show the sentence "<V.3eG> es <V.3eG> <V.3> es <V.3> ich <V.1>" with its morphological variants.
- Semantische Kategorien**: Details show the sentence "<HUM> meinen <PF> <ADJ> <PF> <ADJ+COL+NUM> <N> ((ADJA) | (ADJD)) <PF> ((ADJA) | (ADJD)) & <+COL+NUM> (NN)" with its morphological variants.
- Syntaktische Wortkategorien (extrahiert mit Treetagger von Dr. H. Schmid, CIS)**: Details show the sentence "[NE] sehen [NE] * [NE] [NE] <V> [NE] [VFIN] [NE] [V] [PTKNEG]" with its morphological variants.
- Suche mit Partikelverben**: Details show the sentence "zusammenhängen einfallen herausfallen hervorgehen mitteilen" with its morphological variants.

Spezielle Features von WiTTFind:

Disambiguierung mit lokalen Grammatiken Graphisches Finden (Y. Kalasouskaya)



Spezielle Features

von WITTFind:

Semantisches Finden

über Auswahlmenu:

Forschungsarbeit:

“Das Farbensystem von L. Wittgenstein”

(A. Krey)

The screenshot shows the WITTFind web interface. At the top, there is a search bar labeled 'Semantisches Finden' with a 'Suchbegriff' input field and an 'Übernehmen' button. Below the search bar, there are several sections of semantic classes, each with a checkbox and a list of examples:

- Semantische Klassen für Adjektive und Nomen - Farben**
 - Farben Beispiele für <ADJ> <N>
- Semantische Klassen für Adjektive und Nomen**
 - Numeralia Beispiele für <ADJ> <N>
 - Eigennamen Beispiele für <ADJ> <N>
 - Temporalia <ADJ> <N>
 - Zustände Beispiele für <ADJ> <N>
 - Eigenschaft Beispiele für <ADJ> <N>
 - Ereignisse Beispiele für <ADJ> <N>
- Semantische Klassen nur für Adjektive**
 - Evaluation Beispiele für <ADJ>
 - Relation Beispiele für <ADJ>
 - Komparativa Beispiele für <ADJ>
 - Stilistika Beispiele für <ADJ>
- Semantische Klassen nur für Nomen**
 - Menschen Beispiele für <N>
 - Tiere Beispiele für <N>
 - Pflanzen Beispiele für <N>
 - Objekte Beispiele für <N>
 - Diversa Beispiele für <N>

On the right side of the interface, there are three sections for context options:

- Kontextoptionen**
 - Satzzeichen davor und danach
- Kontext davor**
 - keine Auswahl
 - beliebiges Wort
 - Adjektiv
 - Nomen
 - Verb
- Kontext danach**
 - keine Auswahl
 - beliebiges Wort
 - Adjektiv
 - Nomen
 - Verb

VON DATEN ZU ERKENNTNISSEN I

SEMANTISCHES SUCHEN

semantisches Feld "Grammatik"

neben Einzelwörter können auch Wortphrasen einem Wortfeld zugeordnet und disambiguiert werden.

BISHER:

WITTFind	Grammatik Kalkül Regel Anwendung	Suchen
----------	--	--------

Verstehen von sprachlichen Zeichen <-> Menschen und Menschlichem

Einfache Tests mit syntaktischen Eigenschaften (Info aus Lexikon, Tagging):

WITTFind	<HUM> verstehen	Suchen
----------	-----------------	--------

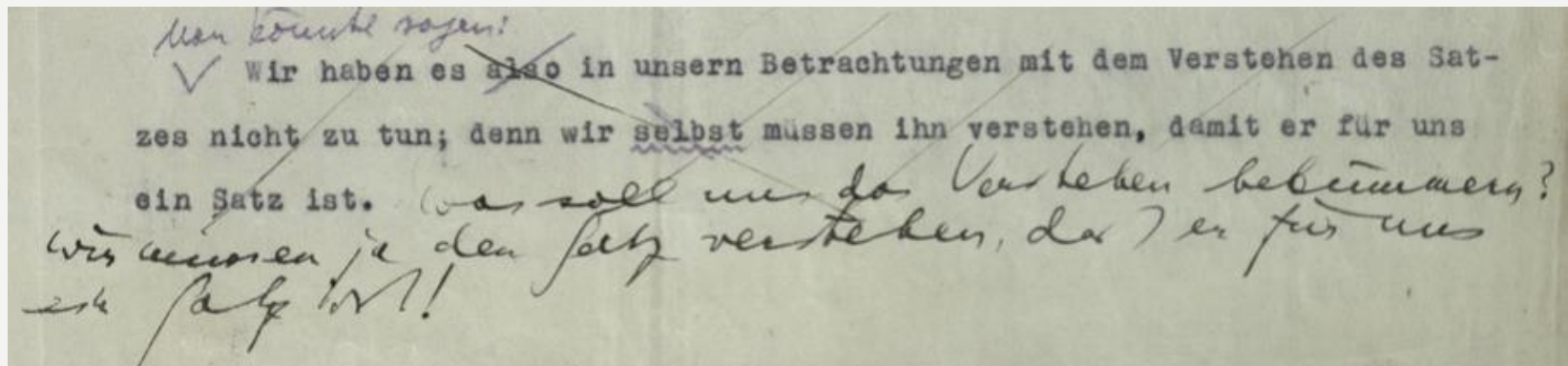
WITTFind	<N> verstehen	Suchen
----------	---------------	--------

VON DATEN ZU ERKENNTNISSEN II

Zusammenarbeit mit "humanities": Wer traut schon einer Edition?

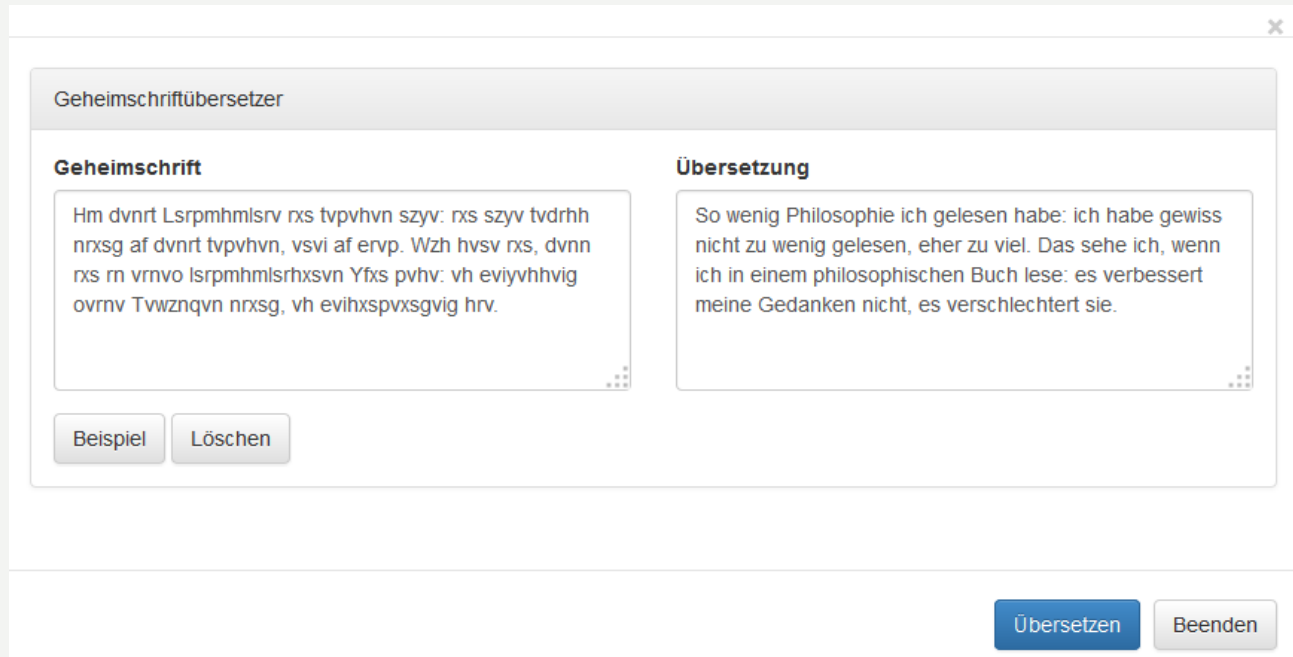
Man könnte sagen: Was soll uns das Verstehen bekümmern?

Wir müssen ja den Satz verstehen, daß er für uns
ein Satz ist!



VON DATEN ZU ERKENNTNISSEN III

Wittgenstein nutzte Geheimschrift... vollständige Datensammlung und
algorithmische Auflösung



Geheimschriftübersetzer

Geheimschrift

Hm dvnrt Lsrpmhmlsr v x s t p v h v n s z y v : r x s s z y v t v d r h h
n r x s g a f d v n r t t p v h v n , v s v i a f e r v p . W z h h v s v r x s , d v n n
r x s m v n v o l s r p m h m l s r h x s v n Y f x s p v h v : v h e v i y v h h v i g
o v r n v T w z n q v n n r x s g , v h e v i h x s p v x s g v i g h r v .

Übersetzung

So wenig Philosophie ich gelesen habe: ich habe gewiss
nicht zu wenig gelesen, eher zu viel. Das sehe ich, wenn
ich in einem philosophischen Buch lese: es verbessert
meine Gedanken nicht, es verschlechtert sie.

Beispiel Löschen

Übersetzen Beenden



VON DATEN ZU ERKENNTNISSEN IV

Finden <-> Suchen

Ludwig Wittgenstein im Ms111,178:

"Wenn ich etwas suche, so ist es wesentlich, daß ich das Finden ebenso ausführlich muß beschreiben können (ob es (je so) eintritt oder nicht) ehe der Gegenstand gefunden ist." Ludwig Wittgenstein im Ms111,178.

Die "humanities" erwarten höchste Treffergüte und Vollständigkeit bei Anfragen
"Nichts darf fehlen" und geringste "Übergenerierung"
(100% Precision, 100% Recall)

Philosoph Prof. Hrachovec (Wien): "Ein neues Rastermikroskop für den Steinbruch"

Leichter Zugriff auf zugrunde liegende Faksimile (Editionsproblematik)

VON ERKENNTNISSEN ZU DATEN

XML zeigt sich als immer ungeeigneter zur Annotation

- Texte sind nicht baumstrukturiert
- XLSX Weiterentwicklung problematisch (Open Source, Michael Kay)
- Texte sind mit Tags überfrachtet

Möglicher Ausweg:

- mehrdimensionale Matrix für unterschiedliche Views
- standoff Markup

DERZEITIGE ENTWICKLUNGEN

Bewertung der Treffer

Weitere semantische Disambiguierungen (Eigennamen, Zeit, Orte)

Masterarbeit zur Geheimschrift Wittgensteins

Ein Quadro- Facsimile- Reader

Statistische Suche mit Solr (M. Wunderlich):

Beta_Version: <http://calculus.cis.lmu.de:4242/solr/#/windex/query>

Automatische Rechtschreibkorrektur (Levenshtein)

Scope Search

Volle Symmetrische Indizierung der

Texte und Autovervollständigung SIS

Daniel Bruder, (see: <http://sis.cis.lmu.de>)



Zusammenfassung: WiTTFind

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

Bearbeitet Dokumentkollektion DM2E texts

Einfache, satzsegmentierte Texte TEI-5 XML (CISWAB),

Multilinguales, elektronisches Vollformenlexikon (WiTTLex),

Wörter, Phrasen, Sätze und semantische Begriffen gesucht im „Zusammenhang des Satzes“

Syntaktische und semantische Disambiguierung mit Lokalen Grammatiken

Darstellung von Faksimileextrakten bei der Trefferanzeige und Faksimile-Reader zum “Blättern”

Lemmatisierte Vorschlagsuche

Leistungsfähige Hilfeseiten und Videotutorial

git- basiertes Programmentwicklung

Browser und Device Unabhängigkeit (Ipad, Handy usw...) (bootstrap),

Betriebssystem unabhängig durch Virtualisierung (docker)

Zitat Ludwig Wittgenstein:

"Wovon man nicht sprechen kann, darüber muss man schweigen." –
Satz 7, Tractatus logico-philosophicus

jedoch:

Die Computerlinguistik kann helfen, die verborgensten Bemerkungen von ihm zu
finden.

(Aber bitte kein neues Rastermikroskop)

Vielen Dank

- für Ihre Aufmerksamkeit
 - dem DM2E EU-Projekt für den AWARD-Preis,
 - an meine Kollegen Alois Pichler und Öyvind Gjesdal,
 - meine Mitarbeiter Daniel Bruder und Florian Fink
- und an meine fleissig programierenden Studierenden:
Ina Arends, Johannes Baiter, Roman Capsamun, Yuliya Kalasouskaya, Matthias Lindinger, Stefan Schweter, Martin Wunderlich und Pascal Zambito