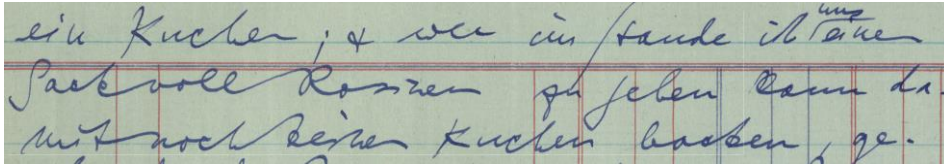


## New (re)search possibilities for Wittgenstein's Nachlass II: Advanced Search, Navigation and Feedback with the FinderApp WITTFind

Hadersbeck, Max; Pichler, Alois; Bruder, Daniel; Schweter, Stefan

Centrum für Informations- und Sprachverarbeitung (CIS), LMU München, maximilian@cis.lmu.de

Wittgenstein Archives at the University of Bergen (WAB), alois.pichler@uib.no



„...wer im Stande ist uns einen Sack voll Rosinen zu geben kann damit noch keinen Kuchen backen...“  
([http://www.wittgensteinsource.org/Ms-136,92a\\_f](http://www.wittgensteinsource.org/Ms-136,92a_f))

### Abstract

We present a new web-based approach to searching and researching Ludwig Wittgenstein's philosophical Nachlass as made available by the Wittgenstein Archives at the University of Bergen (WAB), on Wittgenstein Source (<http://www.wittgensteinsource.org/>). The approach uses highly sophisticated web-technology together with methods and tools from the field of computational linguistics that are developed at the Centrum für Informations- und Sprachverarbeitung (CIS) at the LMU Munich. Tools include the full-form lexicon WITTLex, the “FinderApp” WITTFind, the symmetric autosuggestion tool SIS, a Facsimile Reader with hit-highlighting and an Investigation Mode with an integrated FeedbackApp. The search-methods of the FinderApp include a query language which allows the user to specify exact, lemmatized and grammatical search-queries and a semantic search which permits content driven navigation for colour language and other selected areas. In 2014 our FinderApp WITTFind won the Open Humanity Award within the EU-Project Digitized Manuscripts to Europeana (DM2E).

### 1. “Since last time ...” (Kirchberg 2012)

In this paper we present the second generation of WITTFind, a search engine application that allows users to navigate through the text editions of the Bergen Nachlass Edition (BNE) available on the Wittgenstein Archives' (WAB) Open Access site Wittgenstein Source (<http://www.wittgensteinsource.org/>). This edition gives access in both facsimile and text editions to the Wittgenstein Nachlass, as it was catalogued by G. H. von Wright (von Wright 1986). As of August 2016 the facsimile edition is almost complete, while the text edition so far offers around 5,000 pages in normalized and diplomatic versions. These pages were originally made available in 2009 through the EU financed Discovery project ([http://wab.uib.no/wab\\_discovery.page](http://wab.uib.no/wab_discovery.page)) and include the following Nachlass parts: two items in English from the Bertrand Russell Archives in Ontario, Canada: the “Notes on Logic” manuscripts Ts-201a1 and Ts-201a2 (1913-14), and eighteen items from the Wren Library at Trinity College, Cambridge: the “Lecture on Ethics” manuscript Ms-139a and the “Lecture on Ethics” fair copy typescript Ts-207 (1929), both in English; the volumes (“Bände”) Ms-114 (1932-33) and Ms-115 (1933 and 1936), the notebooks Ms-148, Ms-149, Ms-150, Ms-152, Ms-153a, Ms-153b, Ms-154, Ms-155, Ms-156a (1931-36), the loose sheets Ms-140 (page 39v, 1936) and Ms-141 (ca. 1935) and the typescripts / typescript cuttings Ts-212 (1932), Ts-213 (the “Big Typescript”, 1933),

Ts-310 (the “Brown Book”, 1935) – all except for the Brown Book mostly in German. It is for these Nachlass items, and with them as pilots, that WAB and the Centrum für Informations- und Sprachverarbeitung (CIS) have since 2011 cooperated on developing the advanced FinderApp WiTTFind. In the course of 2016-17, the BNE text editions will be extended to match the facsimile edition, and WiTTFind will be developed further to be applicable also to the additional Nachlass texts. As of today, WiTTFind offers advanced search functions for the 5,000 “Discovery” pages. The search results are displayed in the normalized text edition with parallel highlighting of the result’s locus in the corresponding facsimile. Thanks to the Investigation Mode and FeedbackApp features, WiTTFind additionally offers the capability of studying WAB’s source transcriptions and facsimiles in parallel and, moreover, also to send feedback to the editors on the quality of the transcriptions and facsimiles.

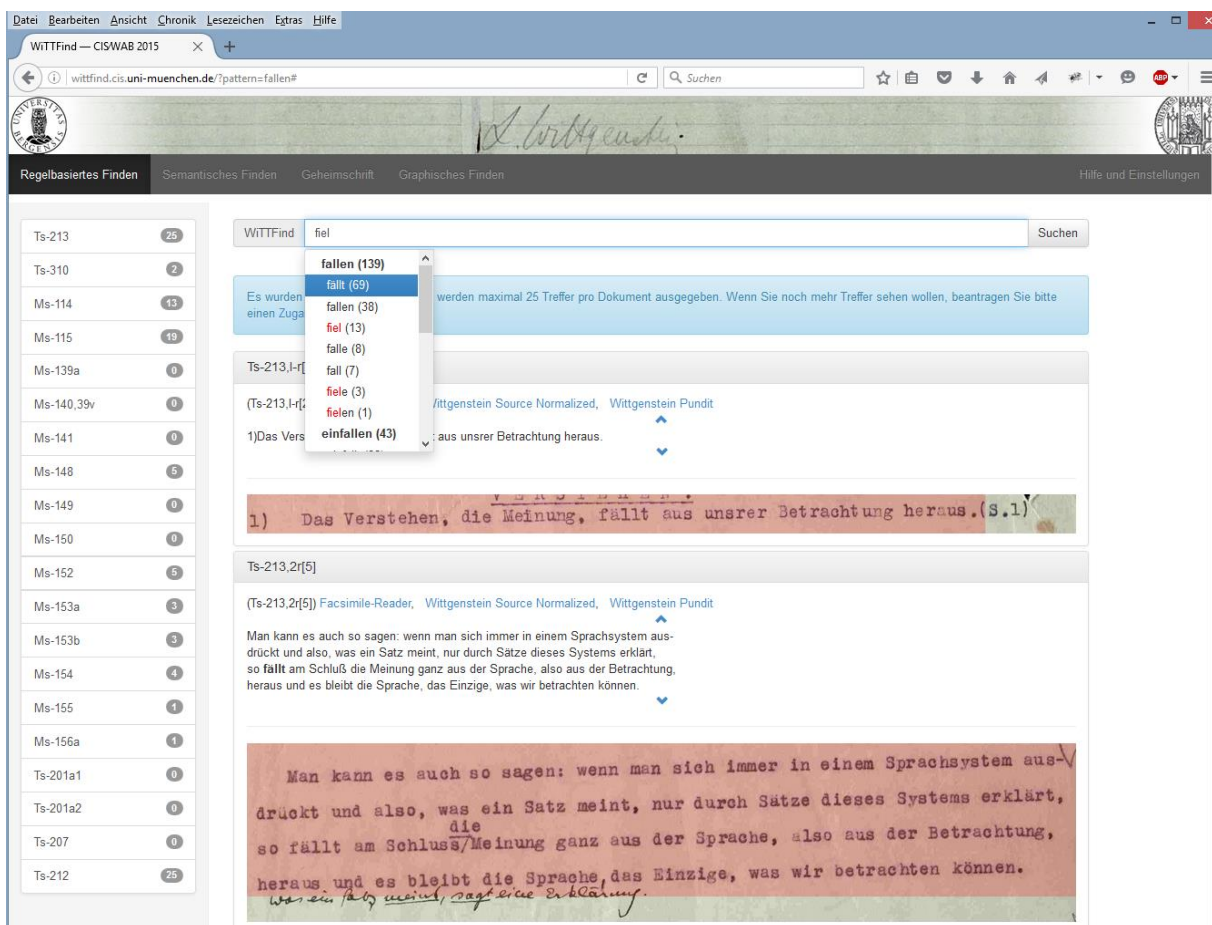


Fig. 1: FinderApp WiTTFind <http://wittfind.cis.uni-muenchen.de>

## 2. Symmetric Autosuggestion, Rule-based Search and Lemmatized Search

Unlike standard (i.e. approximate) search capabilities as e.g. found in Google Books (<https://books.google.com/>) or the Open Library project (<https://openlibrary.org/>), the FinderApp WiTTFind uses specifically tailored rule-based, i.e. “focused”, search-technologies in conjunction with electronic lexica and part-of-speech tagging. Moreover, it provides lemmatized and inverse lemmatized searches. User input is additionally supported by symmetric autosuggestion (Bruder 2012) that is coupled with frequency information on occurrences in the source. When entering more than four characters into the search box the symmetric autosuggestion technology starts: The entered character sequence is expanded to both the left and right to include all the words stored in

the index that feature the same sequence regardless of position (prefix, suffix, and, as a novelty, in infix position). Base lemmas belonging to the word, as well as morphological variants, are suggested and their respective frequencies in the text are subsequently displayed.

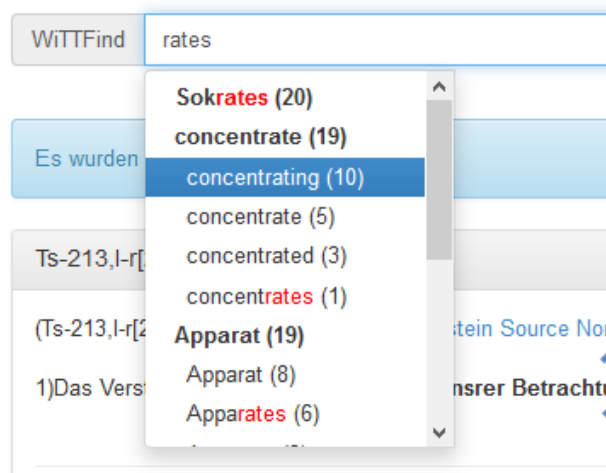


Fig. 2: “rates” expands to “Sok-rates”, “concent-rates”, “Appa-rates” and lemmatizes

We show two examples of lemmatized autosuggestion search for the characters “rates” (Figure 2) and “fiel” (Figure 1): We are offered (frequency numbers in parentheses) the verb lemma “fallen” (139) as well as its morphological variants “fällt” (69), “fallen” (38), “fiel” (13), “falle” (8), “fall” (7), “fiele” (3) and “fielen” (1) – but also “einfallen” (43), “wegfallen” (99) and “mißfällt” (2). Selecting one specific suggestion reveals all occurrences of the specified word in its sentence contexts that are subsequently displayed in the browser with the corresponding “Satzsiglum”<sup>1</sup> with the corresponding facsimile extract highlighted. WiTTFind allows queries which include word form and syntactic as well as semantic and sentence structure specifications (Hadersbeck et al. 2012).

### 3. Facsimile Reader with Investigation Mode and Feedback-App

Clicking on the facsimile extract that is displayed concurrently with the highlighted search hit lets one enter the WiTTFind Facsimile Reader (Lindinger 2015) where the user not only can view the facsimile in question, but thumb and navigate along the found hits through the entire *Nachlass* item. To enable hit highlighting, a semi-automatic tool was developed to extract the geometrical information of remarks with the help of OCR, edition texts and methods of approximate searching (Capsamun 2014). The reader offers a double page view of the source as well as single page views and zoom options for close inspection. Additional hits can be directly accessed within the reader, thus allowing a quick overview of the occurrences in their different contexts. Clicking the zones of specific remarks (“Bemerkungen”) in the reader takes one to Investigation Mode that gives a detailed view of WiTTFind’s XML transcription alongside the facsimile. In Investigation Mode, scholars can not only

<sup>1</sup> The “Satzsiglum” is derived from WAB’s sigla for the single *Nachlass* remarks (“Bemerkungen”). The example from Figure 3 serves as a short explanation: “Ms-115,118[3]\_1” denotes the first sentence in the remark that is composed by the third text block on page 118 of Wittgenstein *Nachlass* item Ms-115. For a more detailed explanation of the sigla system see Pichler 2010: p.164f.

study or compare either the transcription<sup>2</sup> or the facsimile of the source in detail, but also report mistakes, deficiencies and desiderata via the integrated FeedbackApp to the editors at WAB. This feedback functionality is of great use to the editors and can be regarded as a kind of crowdsourcing. At the same time, it also offers benefits to users wanting to learn about transcription techniques and markup. By comparing the facsimile with the XML transcription the user sees which codes (“tags”) were used for recording which writing acts, e.g. <del> for deletions, <add> for insertions, <emph> for emphasis etc. By displaying the use of attributes and attribute values such as <emph rend=“us1”> (for the encoding of passages with one straight underlining) and <emph rend=“uw1”> (for the encoding of passages with a wavy underlining) along with the corresponding facsimile that contains the passages underlined, the tool also offers a small transcription school for aspiring digital philologists, and introduces them to the use of XML elements, attributes and attribute values in the tradition of TEI guided encoding (TEI 2007). The XML transcriptions shown in Investigation Mode can be displayed in different variants: From the open access XML variant at WAB to a reduced XML transcription which is used from the FinderApp. The user fascinated with markup will nevertheless find enough markup and codes accessible in this mode to have his thirst for text encoding matters satisfied, as well as the editors for having their need for correction of the most visible errors.<sup>3</sup>

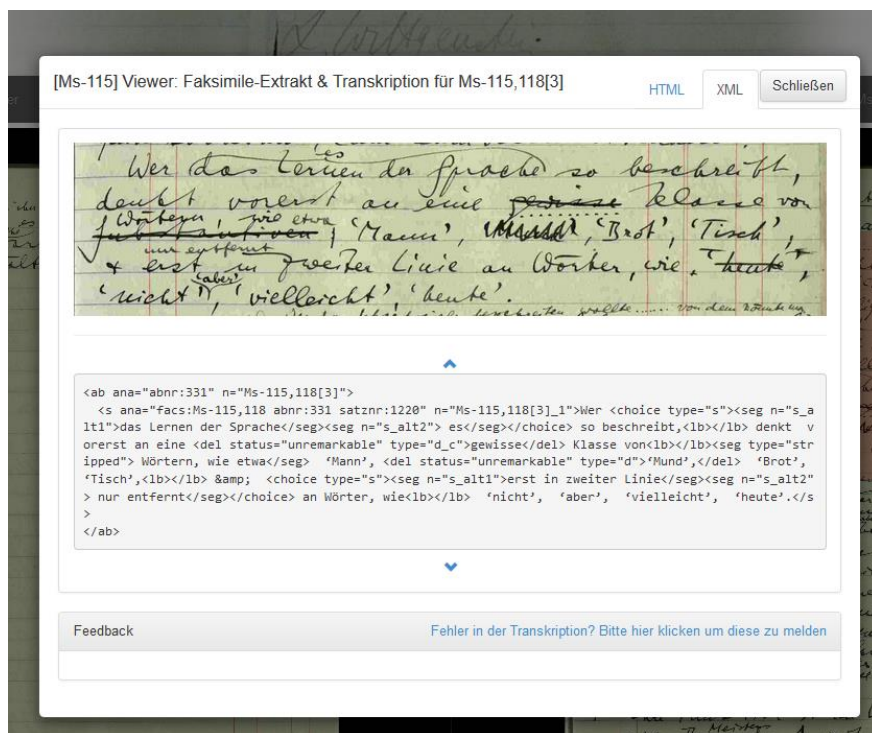


Fig. 3: Investigation Mode and Feedback App

<sup>2</sup> WAB's transcriptions of the Wittgenstein *Nachlass* are marked up in XML. All outputs, whether diplomatic, normalized, metadata or other – including the outputs created for WittFind – are produced via XSLT transformation from the XML transcriptions (on the process, methods and tools see more in Pichler and Bruvik 2014).

<sup>3</sup> A number of alternative text editions as well as “Interactive dynamic presentation” of the WAB text archive of the *Nachlass* can be found at: (<http://wab.uib.no/transform/wab.php?modus=opsjoner>).

```

<ab xml:id="Ms-115,118[3]" xml:lang="de" ana="field:PhilosophyOfLanguage_pub:W-
EPB_date:19360801*-19361130*" emph="blbef_0"><seg type="wabmarks-
secml">&nosectionmark.sm;</seg>
<rs type="extref" key="Augustinus,Aurelius:Confessiones" n="1981:I-8" xml:lang="de"
ana="Biesenbach_AugustinusI0">
<s type="es" rend="indl_2">Wer <choice type="s"><orig type="alt1">das Lernen der Sprache</orig>
<orig type="alt2"><add rend="im">es</add></orig></choice> so beschreibt,<lb/> denkt
vorerst an eine <del type="d_c">gewisse</del> Klasse von<lb/>
<choice type="dsl"><orig type="alt1"><del type="d">Substantiven&colon;</del></orig> <orig
type="alt2"><add rend="i">W&ouml;rter, wie etwa</add></orig></choice>
&usq.susq;Marm&usq.eusq;,<del type="d">&usq.susq;Mund&usq.eusq;,</del>
&usq.susq;Brot&usq.eusq;,&usq.susq;Tisch&usq.eusq;,<lb/> &amp.und;
<choice type="s"><orig type="alt1">erst in zweiter Linie</orig> <orig type="alt2"><add rend="i">nur
entfernt</add></orig></choice> an W&ouml;rter, wie<del type="dnpc">,</del>
&usq.susq;heute&usq.eusq;,</del><lb/>
&usq.susq;nicht&usq.eusq;<corr type="tra">,</corr>
<add rend="im">&usq.susq;aber&usq.eusq;</add>,</del>
&usq.susq;vielleicht&usq.eusq;,</del>
&usq.susq;heute&usq.eusq;&p.es;</s></rs> </ab>

```

Fig. 4 Full XM-transcription for Satzsiglum Ms-115,118[3]\_1

#### 4. Computational Linguistic Tools and Software Development

Our FinderApp WiTTFind is based upon the electronic lexicon *WITTLex* which stores the vocabulary in 58,000 entries with full-form, lemma, morphological, word-form and semantic information. The texts of the 5000 pages were provided by WAB in XML format that along with an XSLT style sheet permits their conversion to readable normalized HTML output. In order to allow for syntactic disambiguation in WITTLex, this output is preprocessed using a state of the art part-of-speech tagger, namely the *treetagger* (Schmid 1995). With the additional help of local grammar techniques, even separated particle-verb constructions like “fällt ... heraus” (from “herausfallen”) ([http://wittgensteinsource.org/Ts-213,1r\[1\]\\_n](http://wittgensteinsource.org/Ts-213,1r[1]_n)) can be disambiguated and thus, unlike classic search engines, ultimately also found (see Fig. 5).

The screenshot shows the WiTTFind search interface. At the top, there is a search bar with the text 'herausfallen' and a 'Suchen' button. Below the search bar, a blue notification bar states 'Es wurden 8 Treffer gefunden'. The main search results area shows a result for 'Ts-213,1-r[2]'. Underneath, there are links for '(Ts-213,1-r[2]) Facsimile-Reader', 'Wittgenstein Source Normalized', and 'Wittgenstein Pundit'. The search result text is '1)Das Verstehen, die Meinung, fällt aus unsrer Betrachtung heraus.' Below this text, there are two small blue arrows pointing up and down. At the bottom, there is a redacted image showing a snippet of the original text: '1) Das Verstehen, die Meinung, fällt aus unsrer Betrachtung heraus.(S.1)'

Fig. 5: Particle Verb separation and disambiguation



The symmetric autosuggestion feature, another tool within our FinderApp, implements the novel index structure *SIS – Symmetric Index Structures* (cf. for a reference implementation and proof of concept <http://sis.cis.lmu.de>), a highly efficient implementation of *Symmetric Directed Acyclic Word Graphs* (Gerdjikov 2012). All our software development around WITTFind is carried out using professional open-source software development best practices including git revision control together with test driven development (TDD), continuous integration (CI) and integrated build system (gitlabci).

## 5. Guided Navigation and Semantic Search

To enable users to access the *Nachlass* through content rather than string, word or phrase search only (Falch et al. 2013), WITTFind has commenced to provide facilities for semantic category based searches. This includes searching for adjectives, numerals, proper nouns and temporalia. Special attention was paid to the handling of adjectives. To this end, eleven groups out of around 7000 adjectives were annotated semantically and together with stylistic information stored in WITTLex (see Fig. 6). As a result, the user can browse through semantical classes of adjectives and follow channels of stylistic “flavour”. Among the adjectives, special attention was again given to the word field of colours (Krey 2014). Here, the subcategories *Grundfarbe*, *Zwischenfarbe*, *Transparenz*, *Glanz*, and *Farbigkeit* were applied. That Wittgenstein himself deals in his Big Typescript ([http://wittgensteinsource.org/Ts-213\\_f](http://wittgensteinsource.org/Ts-213_f)) subchapter “Phänomenologie” ([http://wittgensteinsource.org/Ts-213,436r\[1\]\\_n](http://wittgensteinsource.org/Ts-213,436r[1]_n)) with issues of colour and colour mixture was taken into account in our classifications (see Fig. 7). One lesson from this work was that the application of a standard linguistic classification schemes does not suffice to provide the classifications and tools the Wittgenstein researcher will find useful.<sup>4</sup>

Klasse	Kodierung	Kriterien	Beispiele
Farben	COL	Farben nach Farbenlehre Wittgensteins, Unterkategorie: Grundfarbe, Zwischenfarbe, Transparenz, Glanz, Farbigkeit	rot, klar, blau, gelb, schwarz, rein, klar
Numeralia	NUM	quantitativ	zwei, erst, einzig, viel, weniger, zweite
Relation	REL	„auf ein Wort verweisend oder sich darauf beziehend“/ Fachsprache/ „wie“/ „aus“	physikalisch, wesentlich, kausal, grammatisch, sinnlos
Eigennamen	EN	von Eigennamen abgeleitete Adjektive	Fermatschen, Skolemschen, Sheffersche, Eulerscher
Temporalia	TEMP	Zeitangaben	unmittelbar, ehe, lang, andermal, vorig, meist
Evaluation	Eval	Wertung des Betrachters/ Konnotation/ Subjektivität	falsch, richtig, wahr, einfach, genau
Zustände	ZU	abgeschlossener/ definierbarer Zustand, Lokativa	ganz, wirklich, gegeben, gebraucht, gesagt, allgemein
Komparativa	KOMP	Ähnlichkeit/ Gleichheit/ Steigerung	anders, gleich, verschieden, analog, besser
Stilistika	STIL	autorenspezifische Abstufungen	wohl, wirklich, bestimmt, klar, gewiß
Eigenschaft	EIG	Eigenschaften die Personen zugeschrieben werden	gebildet, ungeschickt, peinlich, selbstständig, gutwillig, naiv
Ereignisse	ER	nicht abgeschlossener Verlauf/ Handlung	unendlich, hinweisend, entsprechend, vorbereitend, folgend

Figure 6: Semantic Classification Adjectives

<sup>4</sup> Complementary work on semantically guided navigation includes the conceptual ontology for Wittgenstein’s philosophy of mathematics and psychology by Addis et al. 2015. For an overall framing of ontological and semantic efforts for the Wittgenstein domain see Pichler and Zöllner-Weber 2013. For a discussion of some issues that emerge in this context see Macha et al. 2013.



Figure 7: Semantic Classification of Colours

## 6. Acknowledgement

The authors would like to acknowledge the help and advice of Dr. Josef Rothhaupt from the Philosophische Fakultät, LMU Munich, and Roman Capsamun, Raphael Höps, Angela Krey and Matthias Lindinger from CIS as well as Øyvind Gjesdal from the University Library at the University of Bergen and Deirdre Catherine P. Smith from WAB for proofreading our paper. We also want to acknowledge the financial support of the DM2E project which contributed to the further development of WITTFind through the Open Humanities Award 2014 (<http://dm2e.eu/open-humanities-awards-round-2-winners-announced/>). The Wittgenstein Nachlass facsimiles are published with the kind permission of: The Master and Fellows of Trinity College, Cambridge; The University of Bergen, Bergen.

## 7. Literature

- Addis, M., Brock, S., Pichler, A. 2015 „Contributions to a Conceptual Ontology for Wittgenstein“ [doi: 10.1515/witt-2015-0111]. In: *Wittgenstein-Studien* 6, 2015, pp. 257-275. Ed. Wilhelm Lütterfelds, Stefan Majetschak, Richard Raatzsch and Wilhelm Vossenkuhl. Berlin / New York: De Gruyter
- Bruder, D. 2012 Symmetric Index Structures -- Highly efficient symmetric text indexing. Master's thesis, CIS
- Capsamun, R, 2014 “WITTFind: Semiautomatische Korrektur des Highlighting im Faksimile”. Bachelorwork, CIS
- Falch, R.J., Erbacher, C., Pichler, A. 2013 “Some observations on developments towards the Semantic Web for Wittgenstein scholarship”. 36th International Wittgenstein Symposium 2013, Kirchberg am Wechsel, *Contributions*, pp. 119-121. Kirchberg am Wechsel: ALWS
- Gerdjikov, S., Mihov, S., Mitankin, P., Schulz, K. 2013. Good parts first-a new algorithm for approximate search in lexica and string databases. arXiv preprint arXiv:1301.0722
- Gross, Maurice 1997 “The Construction of Local Grammars”. In: *Finite-State Language Processing*. Ed. E. Roche and Y. Schabès, pp. 329–354. Cambridge, Mass.: MIT Press

Hadersbeck, M., Pichler, A., Fink, F., Seebauer, P., Strutynska, O. 2012 „New (re)search possibilities for Wittgenstein's Nachlass“. 35th International Wittgenstein Symposium 2012, Kirchberg am Wechsel, *Contributions*, pp. 102-105. Kirchberg am Wechsel: ALWS

Hadersbeck, M., Pichler, A., Fink, F., Gjesdal, Ø. 2014 “Wittgenstein's Nachlass: WiTTFind and Wittgenstein advanced search tools (WAST)”. *Digital Access to Textual Cultural Heritage 2014 (DaTeCH 2014)*, pp. 91-96. Madrid

Hadersbeck, M., Pichler, A., Fink, F., Bruder, D., Arends, I., Baiter, J. 2015 „Wittgensteins Nachlass: Erkenntnisse und Weiterentwicklung der FinderApp WiTTFind“. 2. Tagung Digital Humanities im deutschsprachigen Raum 23.-27.2.2015 (Graz)

Krey, A. 2014 „Semantische Annotation von Adjektiven im Big Typescript von Ludwig Wittgenstein“. Bachelorarbeit, CIS.

Lindinger, M. 2015 „Entwicklung eines WEB-basierten Faksimileviewers mit Highlighting von Suchmaschinen-Treffern und Anzeige der zugehörigen Texte in unterschiedlichen Editionsformaten. Masterthesis, CIS.

Mácha, J., Falch, R.J., Pichler, A. 2013 “Overlapping and competing ontologies in digital humanities” [doi: 10.1145/2517978.2517984]. In: *DH-CASE 2013 Conference Proceedings*. Article No. 5. (USA) New York: ACM International Conference Proceedings Series

Pichler, A. 2010 „Towards the New Bergen Electronic Edition“. In: *Wittgenstein After His Nachlass*. Ed. Nuno Venturinha, pp. 157-172. Houndmills: Palgrave Macmillan

Pichler, A., Zöllner-Weber, A. 2013 “Sharing and debating Wittgenstein by using an ontology” [doi: 10.1093/lilc/fqt049]. In: *Literary and Linguistic Computing*, Vol. 28 (2013) / Number 4. pp. 700-707. Oxford: Oxford University Press

Pichler, A., Bruvik, T.M. 2014 “Digital Critical Editing: Separating Encoding from Presentation”. In: *Digital Critical Editions*. Ed. Daniel Apollon, Claire Bélisle, Philippe Régnier, pp. 179-202. Urbana Champaign: University of Illinois Press

Schmid, H. (1995): Improvements in Part-of-Speech Tagging with an Application to German. *Proceedings of the ACL SIGDAT-Workshop*. Dublin, Ireland.

TEI P5 2007 Guidelines for Electronic Text Encoding and Interchange. Ed. TEI Consortium. 1.4.1, July 1st 2009. <http://www.tei-c.org/Guidelines/P5/> (September 2nd 2009)

Von Wright, G. H. 1986 "Wittgensteins Nachlaß". In: *Wittgenstein* pp.45-76. Translated by Joachim Schulte. Frankfurt am Main: Suhrkamp

[BNE] Wittgenstein, L. 2015- *Wittgenstein Source Bergen Nachlass Edition*. Ed. Wittgenstein Archives at the University of Bergen under the direction of Alois Pichler. In: *Wittgenstein Source (2009-)*. Bergen: WAB