Some observations on developments towards the Semantic Web for Wittgenstein scholarship

Rune J. Falch, Christian Erbacher & Alois Pichler (Wittgenstein Archives, University of Bergen, Norway)

Abstract

The nature of Wittgenstein's *Nachlass* makes it perhaps particularly well-suited for the development of *digital* tools enabling scholars to *navigate* (browse, search, reference) a vast collection of material "criss-cross in every direction" in an informed and efficient manner. (The *Bergen Electronic Edition* of the *Nachlass* is an example of such a tool.) The advantage of an *online* digital platform is being able to do so on different hardware and software systems when- and wherever. And an advantage of *Open Access, Open Source* solutions is that these resources are available for free and the underlying code and datasets open to scrutiny by all (see e.g. Wittgenstein Source and the Agora platform). A valuable addition to current technology will be *lemmatized* searching, enabling search for different forms of terms across both the English and German. Moreover, the development of *semantically* driven tools promises opportunities, yet presents challenges.

1. A possible future scenario?

Imagine a future scholar doing research on, say, Wittgenstein's *Brown Book*. Having read the Blackwell paperback edition thoroughly, filling its margins with pencilled notes she now wishes to systematize them into a study of the notion of "grammar" within the *Brown Book* as well as its relation to the remaining *Nachlass*. She accesses the interactive digital Wittgenstein portal and instructs her computer to run a search for all occurrences of "grammar" and its derivates in the *Brown Book*. Using an integrated tool she annotates the resulting relevant remarks and annotates them with her selected notes. Meanwhile a message pops up on the screen indicating additional information potentially relevant to her: "My information suggests that the *Brown Book*, TS-310, is related to the German MS-115, "Eine philosophische Betrachtung". Would you like me to include results for that as well?" The scholar replies, "Yes, and if applicable, please give me an overview of instances where the one version has an occurrence of the concept of 'grammar' while the other does not in the matching remarks, both from the English to the German and *vice versa*." Making particular note of one of the results, she continues, "I would like to see which notes other scholars have made on this remark." "Would you like to see results both for the primary source and from secondary literature?" the computer asks. "Annotations of the primary source will do for now", she replies. She also decides to narrow down her selection to notes from those who have argued for a contrast between the occurrences in the *Brown Book* and elsewhere in the *Nachlass*. "This should be

plenty for now – I'll deal with the arguments for continuity later on", she says to herself, making a mental note. (Our example is inspired by Szeltner 2013.)

2. Browsing and searching Wittgenstein sources today

While some of this is possible already, the reality of today is admittedly far more mundane and irksome than in the above description. Having said that, the Wittgenstein scholar interested in utilizing digital or online tools already has helpful opportunities at her disposal, and additional ones are not far off.

After more than a decade since its publication the *Bergen Electronic Edition* is still a comparably powerful application. Featuring close to the entire *Nachlass* and a varied set of fine-tuned search functions allowing for detailed access to all 20,000 pages, for the "digital Wittgenstein scholar" it still presents a tool the likes of which are largely unavailable to most other scholars in the humanities. It is undoubtedly a useful application: Consider for instance Peter Hacker's report on his use of it in his revised edition of the commentary volumes on the *Philosophical Investigations* (Hacker 2009, p. xi):

> Now, with the publication of the Bergen edition of the *Nachlass* in electronic form, and the consequent availability of a search engine, I have endeavoured to give all the relevant sources of a given section in the table of sources annexed to each introduction to a chapter of exegesis.

However, the BEE in its current shape does face challenges due to technological development and compatibility issues brought with it. Fortunately, the underlying datasets maintained by the Wittgenstein Archives in Bergen are not proprietary and so lend themselves easily to publication on the web. The next logical step in such a development, therefore, is making these resources available online along with supplementary functionality for powerful navigation (see Pichler 2010).

Indeed, parts of the *Nachlass* were published online on http://www.wittgensteinsource.org in 2009 (see also Pichler & al. 2012 and Pichler & Smith 2013). While it does not (yet) feature the search capabilities of the BEE, it does offer 5,000 pages available in a browser-friendly environment. And while it does not represent a complete edition of Wittgenstein's work, it does have value in making parts of the *Nachlass* available Open Access (contrary to both the BEE and the Intelex Past Masters edition). It also represents a fruitful resource for joint collaboration with several other projects: a Munich group around Max Hadersbeck (see Hadersbeck et al. 2012 and http://wittfind.cis.uni-muenchen.de/) develops solutions for "lemmatized" searching; the Agora project (http://www.project-agora.org/) creates online solutions for the publication of secondary (textual and audio-visual) Wittgenstein sources; and in connection with a range of digital library platforms Net7 are developing the online annotation tool Pundit (http://thepund.it/). Moreover, the *Brown Book* was chosen as the data set for the ongoing "Wittgenstein incubator" experiment (http://dm2e.eu/newsletter-march-2013/) within which both philosophers and computer scientists have joined in order to work with tools for online Wittgenstein research. When sufficiently

developed and brought together these efforts will produce a helpful digital research environment for the Wittgenstein scholar.

The scholar in the introductory example wants to study the notion of "grammar" in Wittgenstein's *Nachlass*, the *Brown Book* in particular. Today, the *Brown Book* material is available in its entirety on *Wittgenstein Source* in both its English (TS-310) and German (MS-115, second part) versions. For "our" scholar, one important task will consist in finding the occurrences of the term "grammar" and its derivates. A truncated search for "gramma*" will yield, among results from other parts of the *Nachlass*, all occurrences of "grammar" in TS-310 or "Grammatik" in MS-115 and their different grammatical forms. For this task, then, a standard search function as we know it from e.g. current browsers will do well enough.

Due to the simple fact that "grammar" and its different forms are all covered by "gramma*" in both English and German, a *lemmatized* search for that term would add little; but if she were to search for other, related terms, e.g. "sprechen", a truncated search would leave out relevant results, while a lemmatized search for the same word would produce a list also including inflections like "gesprochen".
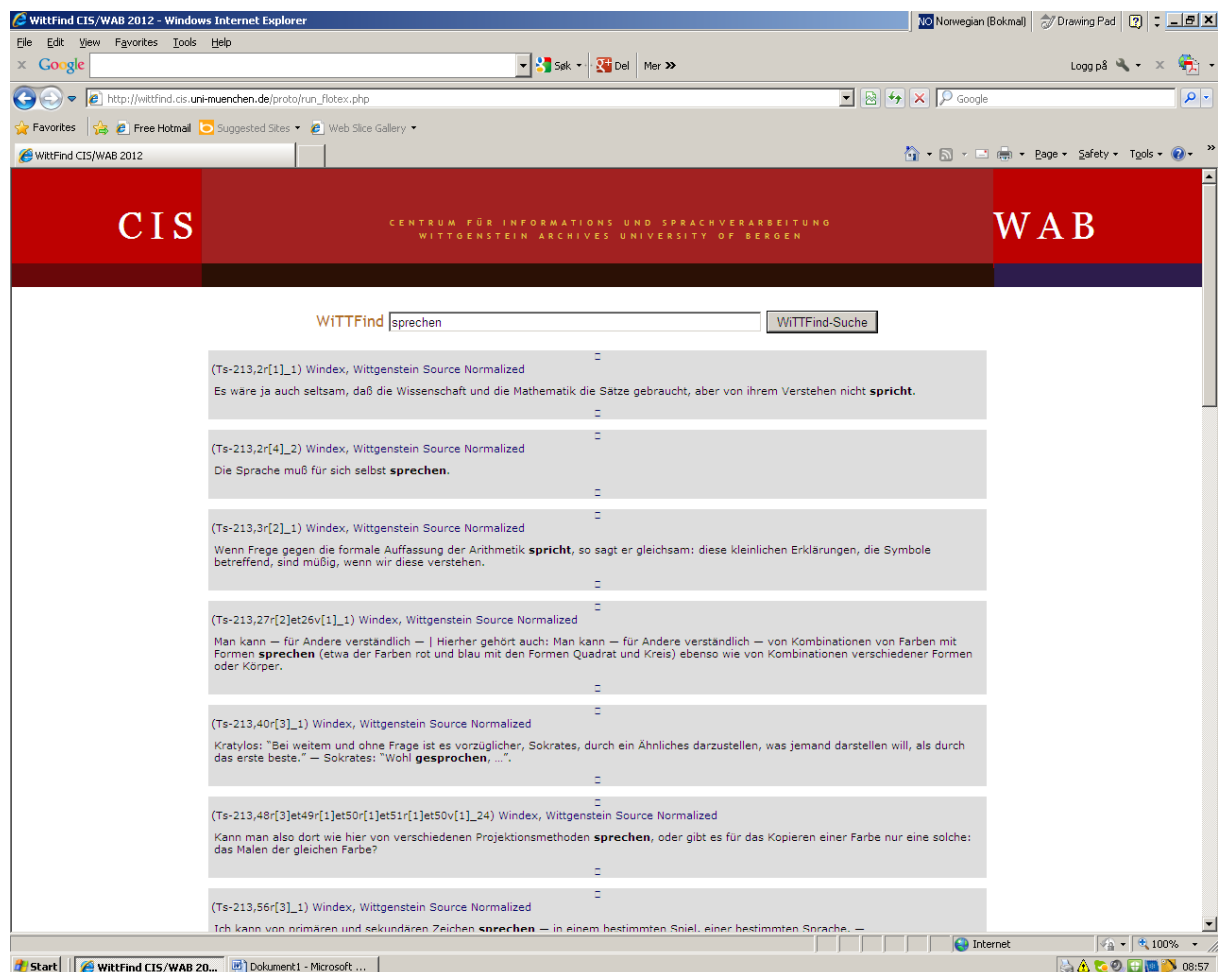

Fig. 1: Screenshot of lemmatized search for "sprechen" in the Big Typescript.

But there is also the issue of delimitation here: We do want all the relevant results, but we might also want to leave out those considered irrelevant; and we might want to be guided towards related concepts. Consider if we were to search for instances of 'thinking' or 'learning'. Here, there would be a number of variants in both English and German, some which would be philosophically relevant, others not. The very first sentence of the *Brown Book* reads like this: "Augustine, in describing his *learning* of language, says that he was *taught* to speak by *learning* the names of things." (http://www.wittgensteinsource.org/Ts-310,1[1]_n; our emphases) Here we have instances of 'learning' which will presumably be philosophically relevant, and we also have an instance of the related concept of 'teaching'. In another remark a bit later on (http://www.wittgensteinsource.org/Ts-310,1[2]et2[1]_n) Wittgenstein introduces the concept of 'training'. In interpreting Wittgenstein these concepts seldom go one without the other. In the last sentence of the first paragraph, "[Augustine] does not primarily *think* of such words as "today"; "not", "but", "perhaps"" (our emphasis), however, we have an instance of 'thinking' which is probably not as relevant for a study of the *philosophical* concept of thinking in Wittgenstein. If all "philosophical instances" of English or German terms expressing central concepts like 'thinking', 'learning', 'understanding', 'teaching' and so on were "tagged" or "marked up" as such, this would represent a navigational function of a *semantic* nature.

Fig. 2: Screenshot showing Wittgenstein Source within the SwickyNotes annotation tool. The contextual graph shows TS-310,147[3]et148[1] selected as the central "node". Here we see several "relations" between the remark and other nodes, among them the concept of thinking.

Semantic search: benefits and challenges

We have proceeded from simple string search to lemmatized search and, ultimately, reached semantic search. A fully-fleshed "Semantic Web" tool would permit the computer to make the sort of suggestions we introduced this paper with. The idea of "semantic" navigation is the upshot of a vision proposed by the founder of the World Wide Web, Tim Berners-Lee (2001). Roughly, the idea is that while the WWW connects web *pages* to each other, a Semantic Web will interconnect (simple) *data* (within or between those pages) at a much more granulated level. Taken to its conclusion, a future Semantic Web would amount to one large database where all data can be connected to each other and thereby queried. (Berners-Lee termed it the "Giant Global Graph" (ibid.) – imagine an unlimited extension of the graph in figure 2 above.) Such interconnection is to be achieved by "tagging" or "marking up" resources (text, pictures, videos, "anything") with additional "meta-data" consisting of information about the resource. Think of this along the lines of a library catalogue index card: the meta-data will include information about who the author is, when it was published, within which field, descriptive keywords, and so on. But moreover and different from an indexing card it will also mark the actual contents with additional information (perhaps a rough analogy could be the way a book looks when it has been read meticulously, highlighted with a marker, the margins scrawled with "see also …", and so on), e.g. that this concept (e.g. 'grammar') or remark is also found somewhere else in the *Nachlass* (e.g. a German version of a remark from the *Brown Book*), etc.

However, although the idea of the Semantic Web is close to 20 years old (see Berners-Lee & al. 2006) its development has been slow, perhaps not surprisinly given the above ambition. They write:

> The Semantic Web is a Web of actionable information—information derived from data through a semantic theory for interpreting the symbols. The semantic theory provides an account of "meaning" in which the logical connection of terms establishes interoperability between systems.

Their talk of "a" or even "the" semantic theory of meaning might give one immediate pause here. Some have claimed the theory alluded to as fundamentally flawed (see e.g. Lemire 2007). Converting all of WWW into GGG graphs would imply a view of all knowledge as something which can be put in a traditional subject – predicate – object form; an ontology of the world mirrored in computational language. However, Berners-Lee's careful stressing of "shared meaning" in a pragmatic manner probably makes the ascription of such a view a bit unfair. Yet reflections on domains within which meaning is often much more vague, ambiguous or contested (e.g. the humanities) than in the fields he considers ("life sciences", businesses) is remarkably absent.

Within the digital humanities some warn against dangers involved in such a vision (see e.g. Erbacher 2011), while others have endeavoured to demonstrate its utility, even for Wittgenstein research (see e.g. Zöllner-Weber & Pichler 2007 and Pichler and Zöllner-Weber 2012). There need be no principal disagreement here. Certainly there are "facts" about Wittgenstein's writings; but many other aspects would be of an interpretive and contentious nature. (Does "grammar" in the *Brown Book* mean "the same" in the *Philosophical Investigations*?)

A semantic tool, involving the application of a "Wittgenstein ontology" to the *Nachlass*, should enable the scholar to do research, not do the research for her; yet it would in one sense decide on some results for her: for while it does open avenues of navigation (e.g. efficient browsing), it also closes some (e.g. search delimitation). Therefore these developmental efforts must constantly be tempered by the realization that many interpretive decisions are involved. A main challenge in developing a semantically enriched *Nachlass* thus lies in the constant deliberation of which aspects should be "marked up" in which way – and of course also involve the wisdom to know when *not* to make such attempts. Just as important, the scholar must be aware of the possibility of being "biased" by the ontology implemented. A way of sidestepping the issue would of course be to switch the ontology off and continue to browse the *Nachlass* without it; or an interesting alternative might be to operate with and compare *several* ontologies, presenting the perspectives of different scholars.

Although the resources currently existing are of a disparate and "R&D" nature, there is hope of bringing them together in the not too distant future. The idea of an online edition of Wittgenstein's *Nachlass* coupled with both lemmatized and semantic search functions and the opportunity of utilizing ontologies to browse and create relations within it holds exciting potential. Such digital tools must on the one hand be sufficiently developed in order to attract scholars; on the other it is clear that we need the research community to take active part in this development, i.e. use and evaluate the tools while they are being developed.The Wittgenstein Archives' own Wittgenstein Ontology can be downloaded from http://wab.uib.no/wab_philospace.page/wittgenstein.owl. We encourage the community to partake and welcome all responses!

Literature

Erbacher, Christian (2011) Unser Denken bleibt gefragt: Web 3.0 und Wittgensteins Nachlass. In Sascha Windholz & Walter Feigl (eds.) *Wissenschaftstheorie, Sprachkritik und Wittgenstein*, pp. 135-146. Heusenstamm: Ontos.

Hacker, Peter (2009) *Wittgenstein : Understanding and Meaning. Volume 1 of an analytical commentary on the Philosophical Investigations*. Co-authored with Gordon P. Baker; second, extensively revised edition by Peter Hacker. Oxford: Wiley-Blackwell.

Hadersbeck, Max, Alois Pichler, Florian Fink, Patrick Seebauer & Olga Strutynska (2012) New (Re)search Possibilities for Wittgenstein's Nachlass, *Contributions of the Austrian Ludwig Wittgenstein Society*, vol. XX, pp. 102-105.

Lemire, Daniel (2007) Yes, the Semantic Web is Flawed, <http://lemire.me/blog/archives/2007/11/02/yes-the-semantic-web-is-flawed/>

Pichler, Alois (2010) Towards the New Bergen Electronic Edition. In Nuno Venturinha (ed.) *Wittgenstein After His Nachlass*, History of Analytic Philosophy series, pp. 157-172. Basingstoke: Palgrave Macmillan.

Pichler, Alois & Amelie Zöllner-Weber (2012) Towards Wittgenstein on the Semantic Web, *Digital Humanities 2012 Conference Abstracts*, pp. 318-321. Hamburg: Hamburg University Press.

Pichler, Alois, Deirdre C.P. Smith, Rune J. Falch & Heinz Wilhelm Krüger (2012) Elements of an e-Platform for Wittgenstein Research, *Contributions of the Austrian Ludwig Wittgenstein Society*, vol. XX, pp. 268-270.

Pichler, Alois & Deirdre C.P. Smith (2013) A list of correspondences between Wittgenstein Ts-310 and Ms-115ii, forthcoming.

Shadbolt, Nigel, Wendy Hall & Tim Berners-Lee (2006) The semantic web revisited, *Intelligent Systems*, *IEEE* 21.3, pp. 96-101.

Szeltner, Sarah (2013) "Grammar" in Wittgenstein's *Brown Book*, forthcoming.

Wittgenstein, Ludwig (1974) *The Blue and Brown Books. Preliminary Studies for the 'Philosophical Investigations'*. Oxford: Blackwell

Wittgenstein, Ludwig (BEE 2000) Wittgenstein's Nachlass: The Bergen Electronic Edition. Wittgenstein Archives at the University of Bergen (eds.). Oxford: Oxford University Press.

Wittgenstein, Ludwig (WS 2009-) Wittgenstein Source <www.wittgensteinsource.org>. Alois Pichler (ed.) in collaboration with H.W. Krüger, A. Lindebjerg, D.C.P. Smith, T.M. Bruvik, V.Olstad. Bergen: WAB

Zöllner-Weber, Amelie & Alois Pichler (2007) Utilizing OWL for Wittgenstein's *Tractatus*. In H. Hrachovec, A. Pichler and J. Wang (eds.), *Papers of the 30th International Ludwig Wittgenstein Symposium*, pp. 248-250. Kirchberg a.W.: ALWS