UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Phil@Scale: Computational Methods within Philosophy

van Wierst, P.; Vrijenhoek, S.; Schlobach, S.; Betti, A.

Publication date
2016
Document Version
Final published version
Published in
Proceedings of the Third Conference on Digital Humanities in Luxembourg with a Special Focus on Reading Historical Sources in the Digital Age

Link to publication

Citation for published version (APA):
van Wierst, P., Vrijenhoek, S., Schlobach, S., & Betti, A. (2016). Phil@Scale: Computational Methods within Philosophy. In L. Wieneke, C. Jones, M. Düring, F. Armaselu, & R. Leboutte (Eds.), *Proceedings of the Third Conference on Digital Humanities in Luxembourg with a Special Focus on Reading Historical Sources in the Digital Age: Luxembourg, Luxembourg, December 5-6, 2013* (CEUR Workshop Proceedings; Vol. 1681). CEUR-WS. http://ceur-ws.org/Vol-1681/Wierst_et_al_Philo_at_scale.pdf

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:08 Nov 2026

Phil@Scale: Computational Methods within Philosophy

Pauline van Wierst*, Sanne Vrijenhoek‡, Stefan Schlobach‡, Arianna Betti†
* Scuola Normale Superiore, Pisa, Italy
p.v.wierst@gmail.com
‡ The Network Institute, VU University Amsterdam
sannevrijenhoek@gmail.com, k.s.schlobach@vu.nl
† The Network Institute / Institute of Logic, Language and Computation, University of Amsterdam (corresponding author)
ariannabetti@gmail.com

Abstract

In this paper we report the results of *Phil@Scale*, a project directed at the development of computational methods for (the history of) philosophy.[1] In this project, philosophers and computer scientists together created *SalVe*, a tool that helps philosophers answering text-based questions. *SalVe* has been tested successfully on the *Wissenschaftslehre* (1837), an extensive work by the Bohemian polymath Bernard Bolzano (1781-1848). Bolzano was a philosopher, mathematician and theologian whose work has been of fundamental importance for the development of Western logic and the foundation of sciences such as mathematics and computer science. The testing of *SalVe* on the *Wissenschaftslehre* reveals that with respect to certain questions within philosophy valuable contributions are obtained by applying even rather simple, well-known computational techniques. We conclude that there is definitely a future for computational methods within text-based philosophical research.

We explain how *SalVe* can be used within philosophical research that relies on textual sources. We will start out with an explanation of our aims in developing *SalVe* and give a short description of *SalVe*'s functionalities, followed by a technical description of the tool. Then we will give a concrete example of how *SalVe* aids philosophical research. We conclude the paper with an evaluation of the potential of Digital Humanities tools for philosophy, and the challenges that face us if we wish to continue this development further.

[1] Phil@Scale was part of the 2012 Academy Assistants project from the VU Network Institute, which is directed at combining expertise from ICT and the social sciences and humanities in order to create innovative technologies. In Phil@Scale, Sanne Vrijenhoek, a master student in Artificial Intelligence, developed the text-mining tool SalVe together with Pauline van Wierst, a master student in Philosophy, under the supervision of Arianna Betti (Philosophy) and Stefan Schlobach (Computer Science). To the project Hein van den Berg, Stefan Roski, Jeroen Smid (all Philosophy), Iris Loeb (Philosophy/Mathematics), Mariya Koleva (Computational Linguistics), Robin Brons (Liberal Arts & Sciences, University College Roosevelt), and Aron de Jong (Lifestyle Informatics) all contributed. See the Phil@Scale project at http://www.networkinstitute.org/academy-assistants/aaprojects/.

Copyright held by the author(s).

1 Our goal

Research in philosophy as it is traditionally done consists for a large part in conceptual analysis and close reading of philosophical texts. Researchers in philosophy come up with hypotheses and confirm or disconfirm them on basis of their analysis and interpretation of the text. This method of research is particularly valuable in the history of philosophy, where researchers read closely very complex, sometimes very long texts that contain concepts far removed from contemporary ones. The method has two main characteristics, which we aim to change for the better. First, slowness, which limits philosophical research to a *small scale*, i.e. research based on a small corpus of a few selected texts. Second, strong reliance on the analysis and interpretation of a researcher working in isolation, which makes philosophical work a *qualitative* and, in some important sense, *subjective* area of research. The subjective aspect of the work of solo researchers in philosophy is mainly due to the fact that the interpretive assumptions forming the background of the research in question are rarely made explicit. Our aim is to put computers to use to make philosophical research faster and larger-scale first, and, second, more quantitative and objective.

It should be stressed that close reading of texts will always be a substantial part of doing philosophical research. It is inconceivable that computational methods will ever completely replace close reading. *SalVe* is therefore developed in order to *facilitate* analysis by philosophy researchers. It does so, for example, by allowing researchers to determine quickly which parts of the textual corpus are most relevant for their research. This connects to our first aim of making philosophical research faster and larger-scale. Furthermore, *SalVe* approaches text as data, and has several options to analyze these data in a quantitative manner. Results from computational analysis can be used to test (and create) hypotheses about texts, and to strengthen the evidence in favor of a particular interpretation of the latter. As such, it connects to our second aim of making philosophical research more quantitative and objective.

SalVe is based on word counts and has several functionalities that deliver information about a textual corpus and its parts. The functionalities are (1) advanced searching for words within a corpus; (2) calculating the similarity between parts of the corpus; (3) displaying the co-occurrence of words (term window); and (4) listing the words that seem most relevant for particular parts of the corpus. Here below we give a technical description of *SalVe*.

2 Technical description

SalVe is a (desktop) Java program based on version 3.6 of Apache Lucene (http://lucene.apache.org/core/), an open source text search engine. *SalVe* analyzes and indexes text files specified by the user. The input text files should preferably be first properly preprocessed, so that they (i) only contain standard characters, and (ii) are split into predefined units (*SalVe*'s 'documents').[2] These units must be determined a priori by the user. They can be single sentences, paragraphs (when the scope is slightly bigger), or even full books when analyzing a large corpus. *SalVe* is also usable on non-preprocessed input, but the quality of the input will influence the quality of the results.

SalVe's core action is *indexing*, during which documents are broken down into *terms*, the position of each term in each document is then determined and efficiently stored in an *index file*. Indexing requires a preprocessing step

[2] We assume here the text files are more than perfectly OCRed copies of the printed originals. By this we mean texts that are not only free of OCR (Optical Character Recognition) errors but also free of any other typographical elements that can interfere with text-mining results, such as end-of-line splits, page numbers, etc. Note that philosophy and other humanities disciplines based on fine-grained textual analysis require exactly this kind of more-than-perfectly-OCRed texts. So, if the text files at disposal aren't of such quality, preprocessing should also aim at improving the files in this sense. Arianna Betti's CLARIN-NL project @PhilosTEI (2013-2014) and follow-ups aim at building workflows for philosophers yielding more-than-perfect TEI-encoded digital copies of printed originals.

Copyright held by the author(s).

relying on *analyzers*, programs that *normalize* the input text. Normalization is a process used to improve comparability of documents. The process transforms texts according to certain norms or standards by reducing certain strings to other strings, e.g. all numerals ('27') to number names ('twenty-seven'). Normalizing a text can mean different things, as the process may be more or less intrusive depending on the purpose at hand. *SalVe*'s users can choose among four different analyzers: Whitespace, Standard, English, and German. The least intrusive analyzer is the Whitespace analyzer, which just *tokenizes* the text, i.e. chops it off into strings enclosed in white spaces. The Standard analyzer also lowercases all words and removes punctuation in addition to tokenizing the text, so that e.g. 'Word', 'word,' 'word.', and 'Word,' are treated as four occurrences of the same term. The Standard analyzer might be useful when dealing with texts in multiple languages. The English and the German analyzers do all the Standard analyzer does, but in addition also filter *stop words* and *stem* the words in the text according to language-specific rules. Stop words filtering removes very common words in a language (such as articles and prepositions), while stemming reduces inflected words to their stem, so that e.g. by using the German analyzer *'analytisch'* and *'analytische'* are treated as two occurrences of the same term - which also means that searching for either will yield the same results. No semantic analysis is applied in *SalVe*.

SalVe calculates similarity between documents by relying on a mathematical representation of textual information based on the so-called Vector Space Model (VSM). In particular, *SalVe* calculate document similarity using the simple cosine similarity metric, which is defined as:

Here below we explain the formula and the basics of the Vector Space Model. First we explain what vectors are.

$$sim(a,b) = \frac{V(a) \cdot V(b)}{\|V(a)\| \cdot \|V(b)\|}$$

A vector is an object that has both a *magnitude* and a *direction*, for example, a force of a certain strength (magnitude) is a vector.

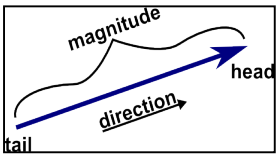


Figure 1: Vector

"Geometrically, we can picture a vector as a directed line segment, whose length is the magnitude of the vector and with an arrow indicating the direction. The direction of the vector is from its tail to its head." (Frank & Nykamp 2015). Vectors can also be represented in a system of coordinates in an n-dimensional space, such as the Cartesian coordinate system in 3-dimensional space. So a 3-dimensional vector is represented as V = (x, y, z) – where the three coordinates (called 'elements' or 'components' of the vector) represent the point assigned to the head, while the origin of the axes is assigned to the tail of the vector. For example, v = (0, 0, 1), a so-called standard unit vector. Multiple vectors can be represented together as a matrix, e.g. this one, where our v = (0, 0, 1) vector occupies the first row.

$$m \text{ (for } m\text{atrix)} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \leftarrow \textit{first row}$$

Copyright held by the author(s).

Recall now *SalVe's* indexing. When *SalVe* creates an index file, it creates in fact a vocabulary, an index of each term in all the documents: it counts how many (different) terms appear in the documents, and assigns them a number. Say you use the English analyzer (which removes stop words) and your collection of documents or *document set* contains the following two (very short!) documents for a total of three terms:

Document a: The cat is on the mat

Document b: The bat is on the cat, and the cat is on the mat

SalVe first creates an index for the terms such that cat = v_1, mat = v_2 and bat = v_3 (there are no more terms to count, since the stop words 'The', 'is', 'on', and 'and' have been removed by the English Analyzer). Then it converts the two-document set into a vector space with a number of dimensions equal to the number of terms in (all) the documents. Since we have three terms in this case, *SalVe* works here with a 3-dimensional vector space. *SalVe* then measures the *term frequency* for each document, i.e. the frequency of a certain term in a certain document. In this way, *SalVe* obtains a 3-dimensional vector for each document (note that since cat = v_1, it corresponds here to the *first* vector element, bat to the second, and mat to the third):

$$v_{\overrightarrow{document\ a}} = (1, 0, 1)$$

$$v_{\overrightarrow{document\ b}} = (2, 1, 1)$$

Indeed, *Document a* has 1 occurrence of cat, 0 of bat and 1 of mat, while *Document b* has 2 occurrences of cat, 1 of bat and 1 of mat. We can also put the two vectors altogether in a matrix:

$$M = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 1 & 1 \end{bmatrix}$$

We can now see how *SalVe* calculates similarity between two documents. Let's go back to our formula:

The formula says: the similarity between document *a* and document *b*, i.e. (*sim* (*a*, *b*)), is equal to the product of

$$sim(a,b) = \frac{V(a) \cdot V(b)}{\|V(a)\| \cdot \|V(b)\|}$$

the vector representation of document *a* (*V*(*a*)) and *b* (*V*(*b*)), divided by the product of the (*Euclidean*) *magnitude* of the vector representation of *a* (||*V*(*a*)||) and *b* (||*V*(*b*)||). To understand what the magnitude of a vector is, recall the representation of vectors as arrows: the (*Euclidean*) *magnitude* of a vector is simply the distance between the vector's tail and its head. Now we're almost done: what the formula calculates, is the difference in *orientation* of the two vectors *a* and *b* (more technically, the formula yields the cosine of the angle between the vectors). By measuring the vectors' orientation instead of their magnitude, we avoid that different document lengths unjustly influence the similarity values. For instance, if document *a* contained cat three-hundred-fifty times and document *b* contained cat two times, on account of the first's being much longer than the second, the *magnitudes* of the vectors would be quite different, but their *orientation* would instead be quite similar. The result of the formula is a value between 0 and 1, where 0 indicates no similarity, and 1 an identical document.

Copyright held by the author(s).

The calculation also incorporates the principle of Term Frequency Inverse Document Frequency (TFIDF or TF/DF). TFIDF divides for each term the frequency in the document by the total number of documents the term appears in. This way terms that only appear in a few documents are more distinctive for that document or cluster of documents than those that appear in a large number of documents.

The results obtained by *SalVe* can be exported to a csv file and used as a basis for further analysis using various programs, such as the graph platform Gephi, which can be used to visualize the similarity relations between terms in documents.

In the following we will show how *SalVe's* functionalities can be used to answer a concrete philosophical research question.[3]

3 A concrete application

To show a concrete application of *SalVe* to philosophical research, we focus on two important notions that are treated in Bolzano's *Wissenschaftslehre*: *analyticity* and *grounding* (see for an extended treatment van Wierst 2013). Roughly described, *analyticity* is the characteristic of a claim thanks to which this claim is *true* (or *false*) independently of the meaning of the terms used in it.[4] *Grounding* (*Abfolge*) is a relation between two (clusters of) claims such that one is the proper scientific explanation of the other.

Analytic truths in Bolzano's view are, for example, truths of the form 'A is A', such as 'A triangle is a triangle'. Bolzano calls all truths that are not analytic *synthetic*. An example of a synthetic truth is 'Triangles have angles that together equal two right angles'. Previous research maintains that Bolzano saw the distinction between analytic and synthetic truths connected with his notion of grounding (which he calls *Abfolge*), although he nowhere explicitly mentions this. The claim is this:

> *De Jong's thesis* According to Bolzano, every *analytic* truth is *grounded* in a *synthetic* truth. (de Jong 2001: 346)

In other words, according to de Jong, Bolzano held that every *analytic* truth stands in the relation of *grounding* to a *synthetic* truth, which comes down to saying that the former is *explained by* the latter. In the following we will show how *SalVe* helps to determine whether de Jong's thesis is correct.

We will start out by asking ourselves what we would have done to confirm or disconfirm de Jong's thesis by philosophical research as it is traditionally done. Close reading and conceptual analysis by a philosopher is often guided by *implicit* research questions. Using *SalVe* requires that we make these research questions *explicit*. In our case, the questions that guide our research would for example be: What exactly did Bolzano write about *analyticity*, *syntheticity* and the *grounding* relation? How do these notions fit in the bigger picture of the *Wissenschaftslehre*? Which evidence is there that Bolzano intended to *link up* the analytic-synthetic distinction with grounding (as de Jong says)?

---

[3] For more introductive information on the vector space model, see http://blog.christianperone.com/?p=2497

[4] Note that, being *analyticity* one of the most debated, elusive and technical concepts in all of philosophy, including its history, this characterization should be seen as capturing only a specific variant of this concept, to wit the Bolzanian notion as it appears in the text we explore. See Rey (2013). Something similar holds for *grounding*.

Copyright held by the author(s).

Now that our research questions are explicit, we reformulate them so that we can use *SalVe's* functionalities to answer them. For example, the question 'What exactly did Bolzano write about analyticity?' can be reformulated as: 'In which paragraphs does the word '*analytisch*' appear most frequently?'; 'Which paragraphs are most similar to the paragraphs in which '*analytisch'* appears frequently?'; 'Which words co-occur with '*analytisch*''? These questions can be answered by means of three of *SalVe*'s functionalities: (1) advanced search for the term '*analytisch*', (2) similarity calculation between the paragraphs in which '*analytisch'* appears frequently and the other paragraphs of the *Wissenschaftslehre*, and (3) the term window of '*analytisch*'.

Advanced search for the terms '*analytisch*' and '*synthetisch*' gives the following result:

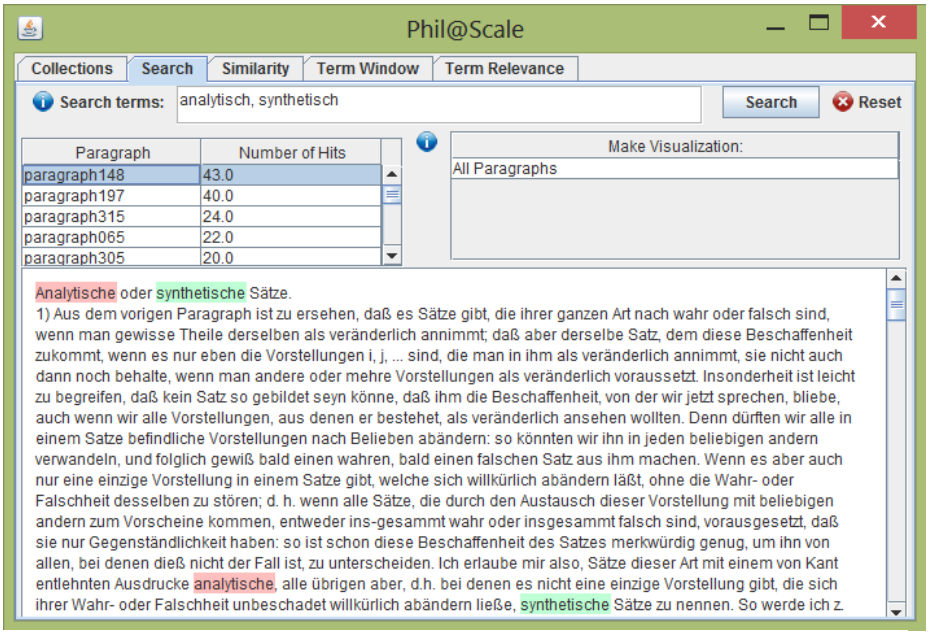


Illustration 1: Advanced search for *analytisch* and *synthetisch*

In the left upper corner, we see a hierarchy of the paragraphs in which these two terms occur most frequently. Only paragraphs are listed in which each of the terms that we searched for occurs at least once. The number of occurrences of the terms we searched for taken together is given in the column 'Number of Hits'. When we click at a paragraph in the left upper corner this paragraph appears in the lower part of the screen, and the terms that we searched for will be highlighted. This allows philosophers to quickly identify the parts of the paragraph where the terms that she is interested in occur, and close read the text around the occurrences of these terms.

Sometimes a philosopher is writing *about* a concept without using his or her technical term (label) for it. This could happen for example when a philosopher writes about a technical concept before the label for it is introduced, or when different labels are used to designate the same concept. In our case, we want to make sure that we do not miss any paragraphs that are *about* analyticity, but in which the *word* 'analytic(ity)' does not occur. Our hypothesis here is that, in comparison to all other paragraphs of the *Wissenschaftslehre,* such paragraphs contain relatively many of the same words as the paragraphs that are about analyticity and in which the term 'analytic(ity)' does occur. *SalVe* has a *similarity calculation* functionality in order to trace these paragraphs. A similarity calculation for paragraph 148, which we identified above as the one in which '*analytisch*' and '*synthetisch*' occur the most, shows that paragraphs 447, 197, and 367, from all the paragraphs in the *Wissenschaftslehre* are the most similar to 148.

Copyright held by the author(s).

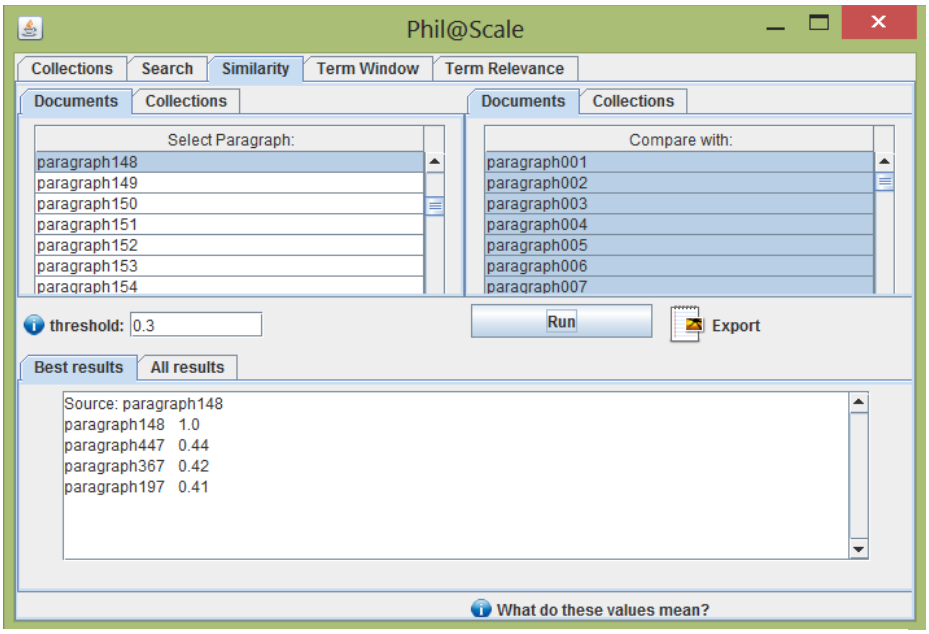


Illustration 2: Similarity calculation between paragraph 148 and the whole *Wissenschaftslehre* with a threshold of 0.3

Note that 447 and 367 did not appear in the top 5 results of the advanced search for '*analytisch*' and '*synthetisch*'. It seems hence that these paragraphs are exactly the ones that we wanted to trace with the similarity calculation: paragraphs in which Bolzano writes a lot about analyticity and syntheticity, without using these terms (or not often). That this application of *SalVe* was successful in this case is shown by the fact that the four paragraphs 148, 197, 367, and 447 are exactly the ones that de Jong found relevant for Bolzano's conception of analyticity (de Jong 2001: 337).

The results of a similarity calculation can be quickly visualized in a program such as *Gephi*. In the following visualization of all the paragraphs of the *Wissenschaftslehre* with each other, we see that the paragraphs that are about analyticity are nicely grouped together. (Recall that paragraph 294 was the 4th most similar paragraph to 148 in the above similarity calculation.) The thicker the line between two dots that represent the paragraphs, the higher the similarity between those paragraphs, and since in this case we set a threshold of 0.3 on the results, paragraphs that do not have a similarity value of 0.3 or higher with another paragraph do not have any lines connecting them with others.

Copyright held by the author(s).

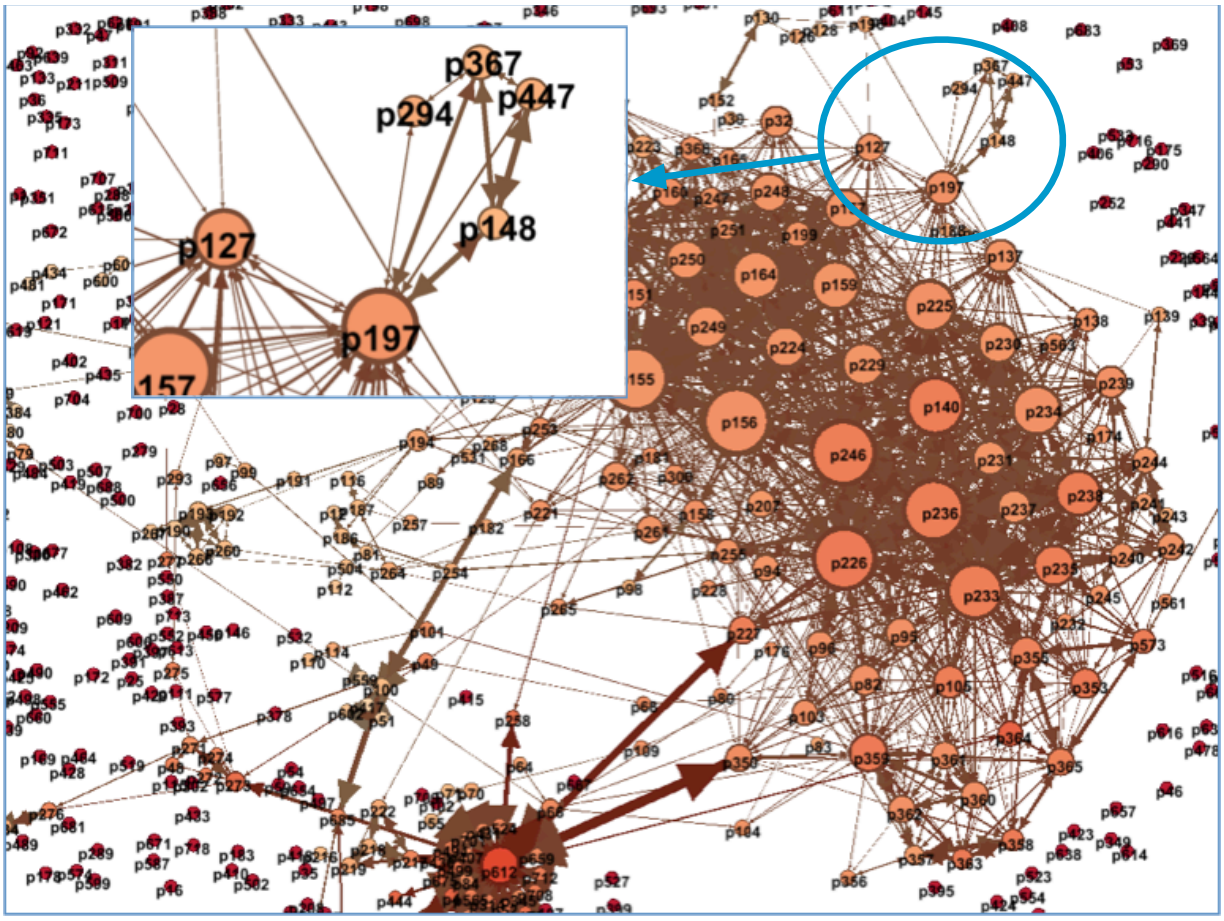
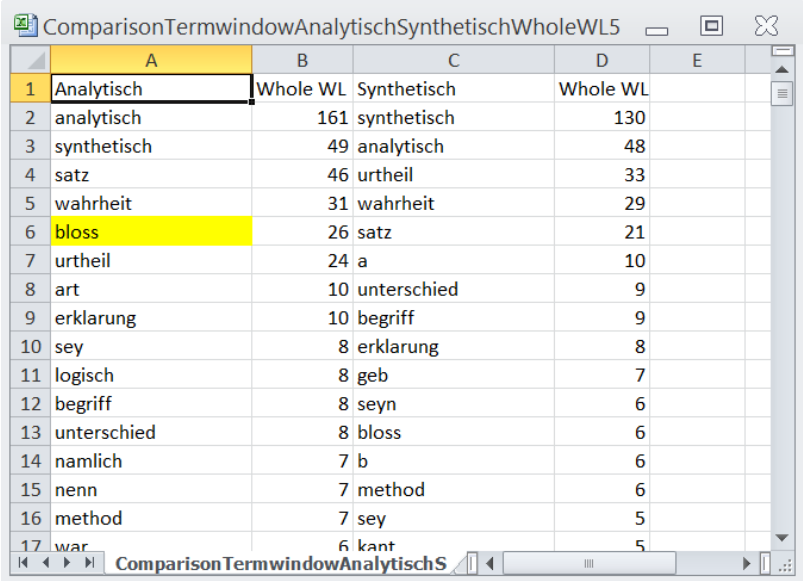


Illustration 3: Gephi visualisation of the similarity between all paragraphs of the *Wissenschaftslehre* with each other. Zoom: Cluster of paragraphs that are about analyticity.

The advanced search and similarity calculation functionalities are useful to determine quickly which paragraphs are useful to read closely. We might also be interested in the number of occurrences of specific terms, and the terms that they are co-occurring with. For this, we can use *SalVe*'s term window functionality. How this can be useful is illustrated below by the comparison of the term windows of *analytisch* and *synthetisch* throughout the *Wissenschaftslehre*.

Copyright held by the author(s).



ComparisonTermwindowAnalytischSynthetischWholeWL5

	A	B	C	D	E
1	Analytisch	Whole WL	Synthetisch	Whole WL	
2	analytisch	161	synthetisch	130	
3	synthetisch	49	analytisch	48	
4	satz	46	urtheil	33	
5	wahrheit	31	wahrheit	29	
6	bloss	26	satz	21	
7	urtheil	24	a	10	
8	art	10	unterschied	9	
9	erklarung	10	begriff	9	
10	sey	8	erklarung	8	
11	logisch	8	geb	7	
12	begriff	8	seyn	6	
13	unterschied	8	bloss	6	
14	namlich	7	b	6	
15	nenn	7	method	6	
16	method	7	sey	5	
17	war	6	kant	5	

ComparisonTermwindowAnalytischS

Illustration 4: Term windows (of 5 words) of *analytisch* and *synthetisch* in the whole *Wissenschaftslehre*

This is the result of searching for *analytisch* respectively *synthetisch* using a term window size 5. This means that *SalVe* counts all words within a distance of five words to the left and to the right of all occurrences of *analytisch* respectively *synthetisch* throughout the *Wissenschaftslehre* (stop words omitted). The comparison of these term windows shows that the word *bloss* ('mere') co-occurs 26 times with *analytisch*, and only 6 times with *synthetisch* (where the co-occurrence of *bloss* with *synthetisch* could also be simply a byproduct of the fact that *analytisch* and *synthetisch* often co-occur with each other, i.e. it might be that one or more of the 6 cases of co-occurrence of *bloss*/*synthetisch* is included within the 26 cases of co-occurrence of *bloss*/*analytisch*). We could see this as an indication that Bolzano saw analytic truths as *less important* than synthetic truths, which in turn would be quantitative evidence in favor of de Jong's thesis that according to Bolzano, every analytic truth is grounded in a synthetic truth. If analytic truths are grounded in synthetic truths namely, then they *depend* on synthetic truths, and in this sense synthetic truths will be more important than analytic ones.

So far, we have a little evidence in favor of de Jong's thesis, and none against it. But in paragraph 315, which we found with the advanced search as the paragraph in which the term *analytisch* appears as the third most frequent, we find Bolzano remarking the following:

(R) Some *wesentliche Lehrsaetze* of pure mathematics are analytic.

He gives an example of such a *wesentliche Lehrsatz*, namely:

(A) $(a + b) + c = a + (b + c)$.

A natural translation of *wesentliche Lehrsaetze* would be 'essential theorems'. In this context one normally thinks of theorems as truths that allow us to prove or explain other truths. For example, (A) allows us to prove:

(E) $(1 + 2) + 3 = 1 + (2 + 3)$.

But according to Bolzano's definition of analyticity (E) is an analytic truth, just as (he says) (A) is: whenever we replace both occurrences of 1 (a) and/or 2 (b) and/or 3 (c) with another number, then the result will be a true statement. So the question is now: is the proof of (E) from (A) in Bolzano's view an *explanation*? In other words,

Copyright held by the author(s).

does (A) *ground* (E)? Because if that is the case, then analytic truth (E) is grounded in analytic truth (A), which means that de Jong's thesis is disproved.[5]

Here it is important to note that, for Bolzano – and this we know from research conducted with traditional methods – the notion of explanatory *proof* reduces to grounding: the grounding chain for a specific claim is an explanatory, i.e. scientific, proof for that claim. We also know that Bolzano distinguishes from grounding another kind of proof, which is not explanatory. Such a proof makes it clear to us *that* something is the case, but does not give us the reason *why*. For example, we can prove that it is warmer in summer than in winter by showing that well-functioning thermometers generally indicate a higher temperature in summer than in winter, but this proof is not an explanation according to Bolzano: it is not the case that it is warmer in summer than in winter *because* these thermometers indicate a higher temperature in summer than in winter (but rather the other way around) (WL §198). As Bolzano sees it, proofs in the second, non-explanatory sense relate truths in a *subjective* order, whereas grounding proofs relate them in the *objective* order.[6] So, the way to go to establish which of the two kinds of proof the proof of (E) from (A) is according to Bolzano, is to figure out whether proofs from *wesentliche Lehrsaetze* in general are grounding proofs. In other words: do *wesentliche Lehrsaetze* occur in the objective order of truths?

In answering this question using *SalVe*, we will make use of the knowledge (again, knowledge that we have from traditional research) that Bolzano discusses the two *ordines* in two different parts of the *Wissenschaftslehre* (de Jong 2001: 329). What we call in the following the *ordo essendi* is the part of the *Wissenschaftslehre* in which Bolzano discusses the objective, or grounding order of truths; what we call the *ordo cognoscendi* is the part in which Bolzano discusses the subjective or non-grounding order of truths. Bolzano uses the term '(*wesentliche*) *Lehrsatz'* in alternation with '(*wesentliche*) *Lehre'*, so in the queries that we did for answering the current question, we search for both terms (see e.g. WL§592).

On the basis of this previous qualitative knowledge obtained with traditional methods, we now make the quantitative step to *SalVe*. In the following line chart, the word *Lehre* is represented as it occurs throughout the *Wissenschaftslehre*. The black line distinguishes the *ordo essendi* (left), from the *ordo cognoscendi* (right). (The utmost left part should be left out of consideration, since this is the introduction and Bolzano discusses both ordines here.)

[5] In this paper we deliberately disregard (and keep for another occasion) an alternative, purely qualitative line of argument based on a suggestion by de Jong to the effect that (A) is, in fact, *not* analytic. We also limit exegetical information and references to the philosophical literature on Bolzano to the absolute minimum needed to expound the results of the *Phil@scale* project.

[6] This is a bit quick. Although *Ableitbarkeit* is, in itself, by no means an epistemic notion relating to knowledge subjects, when Bolzano speaks of objective vs. subjective *order*, he tends to associate 'objectiv' (as things stand) with grounding terminology and 'subjectiv' (as things are known to us) with deducibility terminology. A search for *subjectiv ableitbar* in *SalVe* yields the following two passages: "Insonderheit also, wenn wir ihn für den obersten Grundsatz der ganzen Wissenschaft in objectiver oder nur subjectiver Hinsicht erklären: so müssen wir darthun, daß wirklich alle Wahrheiten, welche nach dem Begriffe unserer Wissenschaft in ihr Gebiet gehören, aus diesem Satze objectiv folgen, oder subjectiv ableitbar sind." (WL §489); "weil sie aus einem und demselben Obersatze entweder nur subjectiv ableitbar sind, oder auch wohl objectiv, d.h. sich wie die Folgen aus ihrem (Theil-) Grunde ergeben." (WL §420).

Copyright held by the author(s).

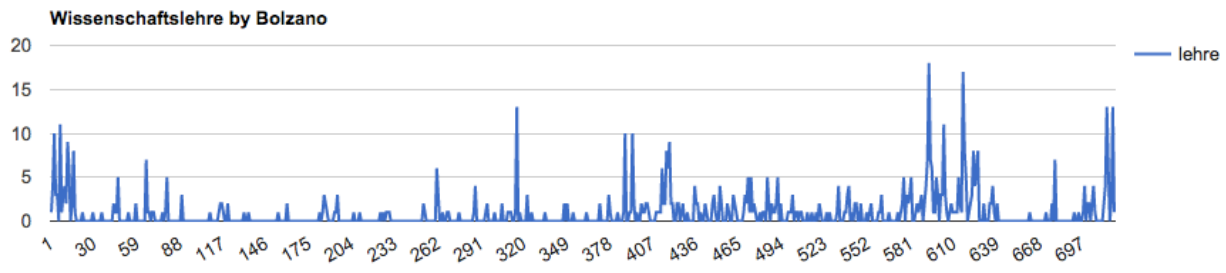


Illustration 5: Line chart of *lehre* throughout the *Wissenschaftslehre*

We immediately see that *Lehre* occurs far more, both absolutely and relatively, in the subjective order of truths. We see the same result when we consider the term window of *Lehre*:

	A	B	C	D
1		Whole WL	Ordo essendi	Ordo cognoscendi
2	lehr	687	153	534
3	wissenschaft	123	18	105
4	uns	101	4	97
5	satz	75	24	51
6	gewiss	60	11	49
7	wahrheit	54	14	40
8	les	54	0	54
9	bloss	50	7	43
10	eben	50	8	42
11	wesentlich	48	5	43
12	muss	47	11	36
13	logik	45	36	9
14	art	44	10	34
15	buch	44	0	44
16	seyn	43	8	35
17	schon	41	10	31
18	vortrag	36	4	32
19	lehrbuch	35	4	31

Illustration 6: Term window of *lehre*

What we see here, furthermore, is that many of the words that co-occur with *Lehre* are terms that are related to the way we come to know truths, hence to the subjective order of truths, such as *gewiss*, *Buch*, and *Lehrbuch*. The same queries for *Lehrsatz* gave similar results, which we cannot all show here out of considerations of space. We will lastly show the occurrences of *Lehrsatz* compared to these of *Abfolge* (i.e. Bolzano's term for the grounding relation) throughout the *Wissenschaftslehre*:

Copyright held by the author(s).

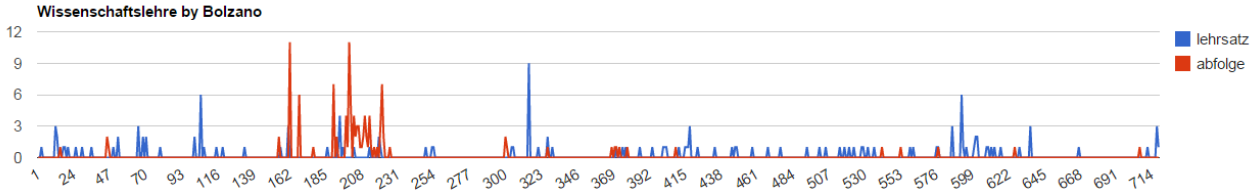


Illustration 7: Line chart of *lehrsatz* and *abfolge*

Clearly, *Lehrsatz* and *Abfolge* have a different distribution: the former occurs more frequently in the parts where Bolzano discusses the *ordo cognoscendi*, whereas the latter occurs more where he discusses the *ordo essendi* (the search functionality revealed that the absolute occurrence of *Lehrsatz* throughout the *Wissenschaftslehre* is 139, and that of *Abfolge* is 113).

We feel justified from this to form the hypothesis that Bolzano saw the role of *wesentliche Lehrsaetze* primarily within the subjective order of truths. And consequently, it seems that Bolzano did *not* hold that the analytic (a + b) + c = a + (b + c) *grounds* (1 + 2) + 3 = 1 + (2 + 3). Bolzano's remark that some *wesentliche Lehrsaetze* are analytic hence does not seem to provide a counterexample to de Jong's claim. All in all, the quantitative information about the *Wissenschaftslehre* that *SalVe* gave us, allowed us to give evidence in favor of de Jong's thesis.

4 Evaluation and future plans

What did the above application of *SalVe* to a concrete philosophical research question show about the use of computational techniques in philosophy? We saw that the advanced search and similarity calculation allowed us to determine quickly which paragraphs of the *Wissenschaftslehre* are relevant with respect to the question that we are posing. This clearly contributes to our aim to make it possible to do philosophical research faster. Besides this, these functionalities point researchers in philosophy to connections between words and parts of the corpus that might not have come to the fore only by means of philosophical research done in the traditional way, that is, only by close reading of the text. In our case we were considering a book of 2500 pages, so it is easily imaginable that some parts of the text that are relevant to our research question escape our attention. Application of *SalVe* increases the chance that we find all relevant passages. Hence, not only does the application of the *SalVe* to philosophy sources contribute to the speed in which philosophical research can be done, but also to the thoroughness of the research.

The term window and (the line chart of) the word occurrences offer quantitative data that count as evidence for a particular hypothesis. In this way *SalVe* contributes to our second aim to make philosophical research more quantitative and objective. In conclusion, *SalVe* offers a welcome complementary improvement to the traditional way in which philosophical research is done. *SalVe* makes it possible to do philosophical research faster and makes this research more quantitative and objective, at least for corpora of a certain size. Importantly, *SalVe* facilitates close reading of the text, and does not restrict in any way what is important for philosophical research. Therefore, the application of *SalVe* to this case shows that computational techniques can be fruitfully applied within philosophy.

We find that it is important for philosophy to develop computational techniques further. One philosophical discipline for which these techniques can offer substantial improvement is the discipline that is called *history of ideas*. This discipline investigates the development of (technical) concepts in philosophical texts throughout history. Ideally, researchers in this area investigate the works of several philosophers through long periods of time, or

Copyright held by the author(s).

compare the works of entire schools.[7] This requires tools apt to proper exploration of massive amounts of machine-readable historical texts in philosophy.

We have experienced that the realization of the latter endeavor comes with several problems that are characteristic for this field. To begin with, at the present time there is little material digitally available. Furthermore, the texts that a researcher in history of ideas is concerned with comprise multiple *languages*, several *scripts*, and are *unstandardized* as to e.g. format, citations, and references.[8] Finally, the concepts that are to be analyzed in this research are of extremely high complexity, which requires particularly challenging semantic analyses. Addressing all these problems in a satisfactory way will be vital to the growth of the emerging field applying computational tools to philosophy.

References

Betti, A. & Van den Berg, H. (2015). Creating a Digital History of Ideas [conditionally accepted]. *Transactions in Digital Humanities*, this volume.

Bolzano, B. (1969/1837). Wissenschaftslehre. In: J. Berg and E. Winter (eds.), *Bernard Bolzano Gesamtausgabe, Reihe 1, Bd. 11-14*. Stuttgart Bad-Cannstatt: Frommann-Holzboog.

Jong, W.R. de (2001). Bernard Bolzano, Analyticity and the Aristotelian Model of Science. *Kant-Studien*, 92 (3), 328–349.

Rey, Georges (2013). The Analytic/Synthetic Distinction. In: *The Stanford Encyclopedia of Philosophy* (Fall 2013 Edition), Edward N. Zalta (ed.), retrieved from http://plato.stanford.edu/archives/fall2013/entries/analytic-synthetic/.

Wierst, P.M.A. van (2013). *Salva Veritate – A master thesis on Bolzanian analyticity and computational methods within philosophical research*, MA thesis, Faculty of Philosophy, VU University Amsterdam.

Biographies of the authors

Arianna Betti is Professor of Philosophy of Language at the University of Amsterdam. After studying historical and systematic aspects of ideas such as *axiom*, *truth* and *fact* (*Against facts*, MIT Press, 2015), she is now trying to trace the development of ideas such as these with computational techniques.

Stefan Schlobach is Associate Professor at VU University Amsterdam. He focuses on using nonstandard techniques and semantics for reasoning and querying, scalable reasoning through approximation and parallelization, and reasoning services for ontology languages such as mapping, explanation, or abduction. He has been involved in several research projects on computational methods in the Humanities, in particular in Philosophy and Social History.

[7] See Betti & Van den Berg (2015) in this very volume.

[8] Currently, solving this problem is at the core of Betti's project @PhilosTEI, see also footnote 2 above.

Copyright held by the author(s).

Sanne Vrijenhoek is currently graduating at the VU University Amsterdam in Artificial Intelligence with an interest in Natural Language Processing. She focuses on applying well-tested techniques from her field of study in the humanities, with the aim of increasing synergy between the two.

Pauline van Wierst graduated in August 2013 at the VU University Amsterdam with an experimental thesis on the notion of analyticity in Bolzano and computational methods within philosophical research under Arianna Betti's supervision. She is currently a graduate student at the Scuola Normale Superiore, Pisa.

Acknowledgements

Work on this paper has been funded by the projects Phil@Scale (van Wierst & Vrijenhoek), Phil@Scale 2 (van Wierst), SalVing GlamMap (Vrijenhoek), ERC Starting Grant Tarski's Revolution #203194 (Betti, van Wierst). Betti has been further supported by the ERC Proof of Concept GlamMap grant #324630 and CLARIN-NL-12-006 project @PhilosTEI. Van Wierst has been further supported by the Prins Bernhard Cultuurfonds, the Fundatie van de Vrijvrouwe van Renswoude te 's Gravenhage and by a DHLU 2013 travel grant that enabled her to travel from US to deliver the talk which is at the basis of this paper.

Copyright held by the author(s).