Digital Ricœur

George H. Taylor and Fernando Nascimento

University of Pittsburgh and Pontifical Catholic University of Campinas

Abstract:

As Ricœur scholars know, the literature by and on Ricœur is vast. Material written by Ricœur that is not collected in published volumes is often difficult to locate, and even in the published volumes it is frequently a challenge to locate where Ricœur discusses a particular topic. Given the amount of his work it can be a challenge too to determine changes in his analyses over the life of his corpus. And locating secondary literature on Ricœur can be equally problematic. In response, we have been working to establish a model for how Ricœur's corpus might be digitized so that the issues of access, keyword location, and pattern might all be addressed. To develop the model, we are starting with Ricœur's primary texts in English and plan to expand over time to other languages and to the secondary literature on Ricœur. In the present article we discuss our model and its five steps: digital access; copyright; text preparation for digital searches and analysis; examples of digital searches and analyses; and an archive portal interface that allows users to query based on an extensible set of search variables without needing to know or access the underlying search logic. We also invite interested researchers to help assist the development of this digitization project.

Keywords: Paul Ricœur, Cultural Hermeneutics, Phenomenology, Social Imaginary, World Horizon, the Symbolic, the Human Condition.
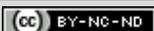
Résumé:

Comme le savent les spécialistes de Ricœur, l'œuvre de Ricœur comme la littérature secondaire qui lui est consacrée sont vastes. Les écrits de Ricœur qui ne sont pas réunis dans des volumes publiés sont souvent difficiles à localiser, et même dans les volumes publiés, c'est fréquemment un défi que de localiser dans quels écrits il aborde tel ou tel thème. Étant donné l'ampleur de son œuvre, determiner quels ont été les changement de ses analyses tout au long de son corpus peut aussi constituer un défi. De même, localiser la littérature secondaire sur Ricœur peut s'avérer problématique. En réponse à cela, nous avons travaillé à établir un modèle de numérisation du corpus de Ricœur, de façon à ce que les questions d'accès à l'œuvre, de localisation par mots-clés et de modélisation puissent être toutes solutionnées. Pour developper ce modèle, nous commençons avec les écrits de Ricœur en anglais et nous prévoyons avec le temps d'étendre cette demarche aux écrits du philosophe dans d'autres langues ainsi qu'à la littérature secondaire le concernant. Dans cet article, nous présentons notre modèle dans ses cinq étapes: accès numérique; droits

d'auteurs; préparation de textes pour des recherches et des analyses numériques; exemples de recherches et d'analyses numériques; interface de portail destiné aux archives et basée sur un ensemble extensible de variables permettant aux utilisateurs de faire leur recherche sans avoir besoin de connaître ou d'avoir accès à la logique de recherche sous-jacente. Nous invitons également les chercheurs intéressés à contribuer à ce projet de numérisation.

Mots-clés: Paul Ricœur, herméneutique culturelle, phénoménologie, imaginaire social, l'horizon du monde, la symbolique, l'humaine condition.

Digital Ricœur

George H. Taylor and Fernando Nascimento
University of Pittsburgh and Pontifical Catholic University of Campinas

As Ricœur scholars know, the literature by and on Ricœur is vast. In 2008 a primary and secondary bibliography on Ricœur was published in book form that was itself nearly 600 pages long.[1] The volume included every article and book published by on Ricœur in every language where publication has occurred, some 32 languages. It can be difficult to master Ricœur's own corpus as published in one's home language let alone in material not yet translated, and with the international growth of Ricœur studies, it is becoming increasingly difficult to read all the secondary work on Ricœur too, even within any individual's more narrow areas of scholarly interest. It can also be difficult to locate material, even if we know it exists. A number of Ricœur's articles in English were published in obscure journals to which many university libraries do not have subscriptions, and many Ricœur articles in French were published in smaller journals that are also very difficult to locate and have not been collected in larger Ricœur volumes. Locating secondary literature on Ricœur can be equally problematic. The problems of access are exacerbated for the many Ricœur scholars internationally who are either not affiliated with universities – or their libraries – or they are attached to universities whose library budgets for items such as periodical subscriptions are quite limited.

Also a problem, even if we have access to Ricœur's work, it can often be difficult to figure out where he has addressed a particular thinker or theme, given not only the volume of his work but the way he interweaves topics and subjects in the midst of larger theses. Where, exactly, does Ricœur discuss the work of Hannah Arendt, Ludwig Wittgenstein, or Enrique Dussel? Where does he discuss Buddhism? At several points in his corpus he discusses the topic of a "practical concept," but does he ever define what he means by this term?

An additional problem is that it is now a challenge to determine patterns or breaks in Ricœur's corpus more generally. How does his view of hermeneutics change over time? In what ways do the scholars he engages with vary over the years? What are his responses – positive and critical – to Hegel across his texts? Currently, only very laborious readings by individual scholars address these kinds of issues.

How might we better address the problems of access, keyword location, and patterns in Ricœur? Historically, many thinkers would have collected volumes of their work published, and the volumes would have elaborate indexes. Today publication of these collected volumes is an expensive proposition, and it would still limit scholarly access to libraries that could afford these volumes. And access would be an additional difficulty across languages, as the volumes would be less likely to be purchased by a library if written in a foreign language. Further, even with access, determining the location of references would depend on the adequacy of indexes and would not solve the question of the patterns and breaks in Ricœur's work.

In response, we have been working to establish a model for how Ricœur's corpus might be digitized so that the issues of access, keyword location, and pattern might all be addressed. To develop the model, we are starting with Ricœur's primary texts in English, because that is the largest amount of material after the work in French. We are leaving the material in French for the Fonds Ricœur to digitize, which it has been doing, and we are coordinating our digitization effort with them. In this essay we discuss our model for the digitization of Ricœur's primary texts in English. Once this model is thoroughly tested, we hope to expand to primary material in other languages and then expand to secondary material in various languages. All of the digitized material would be accessed by links that would be created on various Ricœur-related websites.

Our digitization project includes five elements: digital access; copyright; text preparation for digital searches and analysis; the digital searches and analyses; and an archive portal interface that allows all users to query based on an extensible set of search variables without needing to know or access the underlying search logic. In what follows we discuss each of these stages.

1. Digital Access

An initial issue that we have been working on for a few years now is how to make the primary material on Ricœur more readily available, including internationally. We have begun initial discussions with the Fonds Ricœur about making Ricœur's work accessible digitally, so access could be online. We are happy to share with the Fonds Ricœur anything in our digitization model that might be of interest to them. Digital access would also allow for digital searching, which would be essential for keyword and pattern analysis.[2]

The provision of access to the primary digitized materials would in itself be a significant advance. Following the presentation of our project at several international Ricœur conferences over the past 18 months, we received a number of audience comments endorsing the particular value of international access to Ricœur documents that are very difficult to locate particularly outside of their countries of origin.

Consistent with our current model, thus far we have been particularly gathering digitized primary Ricœur texts in English. Many of these texts are PDF copies of material that we lawfully already owned or downloaded from our university subscription services. Other copies were PDFs of Ricœur volumes supplied to us by other Ricœur scholars. Because we owned the original hard copies of these books, current U.S. law allows us to engage in "space shifting" – the transformation of media from one medium to another, here from hard copy to PDF – even though the PDF did not come from our own hard copy. As a matter of "fair use," it is not a copyright violation for us to use these gathered texts for own scholarship.[3] We return to the question of fair use of these materials by other scholars in part 2.

With the aid of our research assistant, Krishna Hegde, we have established a database of the citations to all Ricœur books and articles in English, a citation list that we have incorporated from the published bibliography. This database is allowing us to track the items that we have in initial PDF form; their conversion into plain text; the review of the plain text against the original to create a "clean" plain text; and the incorporation of each plain text into our DH analysis. These procedures will be discussed in more detail in part 3. Because we currently are focused on establishment of proof of concept of these procedures, we are dedicating less time and attention

to ensuring that we have obtained all of Ricœur's publications in English for the database. The limitation in the completeness of the database will of course affect any search analyses we run at present, but those will be corrected as we complete the database over time.

2. Copyright

As noted in part 1, many Ricœur scholars internationally do not have access currently to portions of Ricœur's articles and books, whether in their home language or in French or English. It would be optimal, then, for scholars to have access to these materials in full text at a digital Ricœur archive for both initial reading and then analysis. Provision of these materials in full text on a digital archive would, however, violate the texts' copyright protections unless we receive permission from the copyright holders for the items' inclusion. A second element of the larger digitization project, then, is the issue of copyright. We would need the permission of Ricœur's estate in order to create a full text digital archive. We would also need the permission of existing publishers. Preliminary discussions with the Fonds Ricœur on the copyright issue are underway. Again with the help of research assistants, we have compiled a database of Ricœur material in English by publisher so that we can write to seek publishers' permission to include the full texts they control in the archive. We have also created a database of secondary material on Ricœur in English by author so that we can ultimately write these authors for similar permission to include their texts as well. The creation of these databases has been a major undertaking in itself. Similar efforts would need to be undertaken in other languages. As is well known, many journal publishers have granted libraries licenses to provide their materials online, but these libraries have had to pay to subscribe for this online access. We do not know to what extent publishers would seek to require payments for licenses for Ricœur material also or allow free access. We have greater hopes for obtaining publisher permission to grant the right for inclusion of full text articles rather than books, but would write all publishers. Some of Ricœur's books in English are out of print, so they may present greater possibilities for publisher permission to include them.

For quite practical reasons, we have as yet not written any publishers to seek permission for full text access on the archive. We think our efforts are better spent digitizing the primary materials, both because the writing of publishers would require considerable time and, under recent U.S. legal rulings, we can allow researchers to undertake digital searches of the material on the archive if, as we go on to detail, we restrict their results to relevant "snippets." In June 2014 a U.S. Court of Appeals held in the *HathiTrust* case that it is permissible for a digital archive to digitize the full text of a copyrighted article or volume to permit a user search as long as the results of that search provide only a summary of the data sought, such as page numbers for the search term and the frequency with which the term appeared on any text page.[4] No portion of the volume's text itself is shown. It is then not a copyright violation on its own for a library or archive to digitize a copyrighted work to permit full-text searches. More recently, in October 2015, the same appellate court held in the case of *Authors Guild v. Google* that it was not a copyright violation for Google to digitize books under copyright, allow researchers to conduct electronic searches of the text, and provide the researchers access to relevant snippets of the text.[5] Importantly here, the court's holding also rested on the fact that Google permitted access to only a restricted number of snippets of relevant text and so not every reference.[6]

The major significance of these rulings for our project is that it appears to be accepted under U.S. law to date that an archive can create a full text database for researchers to search, even if the limitations on what results of the search the archive can provide remain somewhat ambiguous. At the very least, we can create a full text database of all Ricœur materials and provide researchers access to it (for limited purposes), even if the researchers do not own the materials themselves in hard copy. An entire Ricœur text could be loaded onto the digital archive, then, without the publisher's permission, and as long as we provide researchers only limited access to the text in accord with their search, then that is not a copyright violation. This "fair use" of a copyrighted text does not violate the copyright. Fair use promotes the growth of scholarly knowledge by building on and referencing the work of those who have gone before.

3. Preparation of Texts for Digital Analysis

On the basis of parts 1 and 2, we have been gathering Ricœur texts and know that we have the right to create a full text database from them. In part 3 we review briefly the steps required for preparing the texts for digital analysis. Almost all of the Ricœur texts we have gathered are PDFs. (A few are copies of unpublished manuscripts in Word in English. We focus here on the PDFs only.) Initially we must determine whether the PDF text has a layer that permits conversion to plain text. If not, we must use Optical Character Recognition (OCR) software to convert the PDF to plain text. For most of our texts, conversion to plain text has not been an issue. For those texts that require OCR, we have used Adobe Acrobat OCR, for which we have a license. Once we have the plain text, we then have used the Stanford Natural Language Processing (NLP) software version in Java to create the NLP tagged version of the text. NLP is particularly useful for tagging parts of speech in a text, and as we describe in part 4, on the basis of NLP we have run some sample experiments on patterns in Ricœur's work.

We have begun the laborious process of comparing the NLP tagged text to the original PDF to ensure that we have a "clean" text that will generate accurate statistics for the kinds of more complex digital searches we discuss in the next part. To establish proof of concept we initially restricted this review to one individual, our research assistant Krishna Hegde, so that on the basis of his work we could better identify what textual elements are in error (e.g., punctuation marks, spaces between words) and what elements should or should not be excluded (e.g., page headings, captions, section headings). Now that our method of review is established, *we invite other Ricœur scholar volunteers to participate in this review of other primary texts*. (Anyone interested should email co-author Fernando Nascimento (fern.nascimento@gmail.com).) The ability to draw upon these volunteers will make this phase of the research much less burdensome than if we had to undertake the task ourselves for our individual research.

We have not waited to establish clean texts in order to run our experiments and are aware that at points our results are infected by text errors. As described below, in one of our Word Clouds, for instance, the term "iiii" appears. These errors do not seem to detract from the major insights of the experiments, and they will be eliminated as the process of cleaning the plain texts continues.

4. Digital Searches and Analysis

Our main focus in this part is on offering examples of the more complex computer searches that will be available with access to the digital primary material, but we begin with the broader value for researchers of digital searches of Ricœur's corpus. As we have mentioned, some of the material available for research through a digital database may have been previously unavailable to a number of researchers, particularly those coming from countries whose home language – and bibliographic resources – are not in English. It will also be extremely valuable for researchers to have one database to explore rather than need to work through independent searches of multiple, separate texts located in different databases even if available electronically. And an electronic database will of course allow for simple searches of an article, volume, or multiple texts that previously would have required the researcher's time-consuming visual inspection. These simple electronic searches may prove extremely illuminating for any researcher but are particularly fruitful in Ricœur studies. As scholars are aware, Ricœur is particularly well known both for the breadth of his own scholarship and for his ability to drop in nuggets of information on a specific subject when the ostensive theme of the article or volume is something quite else. He will write something quite intriguing about Kierkegaard, for instance, in an article on Hegel. He will tantalize with comments here and there about the term "practical concept" without providing an explicit discussion or definition. Searches across his texts for his portrayal of Kierkegaard, the practical concept, or Wittgenstein, Arendt, or Buddhism will prove quite fascinating to Ricœur researchers. While these kinds of searches are fairly simple in terms of computer search logic, their results may be potentially quite far-reaching.

Part of the larger point here is that we are mindful that there is no necessary correlation between the sophistication of a search logic procedure and the substantive value of the results. A simple search may prove very insightful, and a sophisticated search may amaze in its complexity yet leave us shrugging our shoulders at the lack of depth in the results. We have tried to remain attentive to the lesson of substantiality in appraising our more complex experiments, to which we shortly turn.

As a final introductory point to this part, our experiments – and conference audience response to them – have required us to reconsider some of the framework for our experiments that we did not initially consider. A considerable portion of the issue here arises from the fact that most of Ricœur's texts in English are translations of work published previously in French. When we first ran our experiments, we used the dates of English publication to plot Ricœur's usage over time. As pointed out, though, by one conference participant, some of our results were skewed because of the discrepancy between the year of original publication in French and the year of publication in English. We had not given that problem any real attention. We have had to expand our database listing Ricœur's work in English to include dates of original publication in French and have reworked experiments in light of this change in dates. (The original results, which we do not include here, might still be pertinent to the reception of Ricœur in English by his readers.) We will also need to disaggregate volumes of collected articles by the year of initial publication of each article.

Another challenge presented by the Ricœur translations is that the translations of his vocabulary may not be uniform across translators (e.g, *objectivation* translated as *objectification* or

objectivation; *distanciation* as *distantiation* or *distanciation*), and meaning indicated by one term in French (e.g., *sens*) may be indicated by multiple terms in English *(meaning, sense)*. These problems obviously affect searches involving word counts. One of our Word Clouds, for instance, shows the two different spellings of *distanciation* and *distantiation*, and that affects the significance of that term in relation to the more prominent terms of *language*, *text*, and *translation*.

Also a problem for word counts is the appearance of the same Ricœur article in English in several collected volumes, in addition to its initial appearance in a periodical. A word count running across his corpus would lead to a multiplication of citations from the same article. And for communicating searches involving page references there is a similar problem of registering those references when other researchers may possess or refer to page references in publications elsewhere. This is an issue for both Ricœur articles and some of his books. His book, *The Rule of Metaphor*, for instance, was published over time by two different publishers in editions with different paginations. We have not yet resolved these difficulties.

We now turn to our development of more complex computer experiments to demonstrate the potential new insights that may be generated by more sophisticated computer analysis of Ricœur's digital corpus. As we explain in part 5, we will be developing a portal to the digital database that will allow any researcher to undertake this kind of research without further training. Our research into these more complex forms of computer analysis has been assisted by our engagement with the tools of the *digital humanities*. The digital humanities is a subject that has been around for a few decades but has more recently gained greater general scholarly attention. The question is how can computer search techniques provide different and perhaps more revealing forms of scholarly insight in the humanities. While in the humanities we typically think of interpretation according to a "close reading" of the text, where we analyze the text word by word, line by line, and paragraph by paragraph, the digital humanities offers the possibility of a "distant reading,"[7] where we can assess patterns or discontinuities across multiple texts. These tasks would be extremely complicated and time-consuming to undertake by us individually; use of computer analysis deepens and broadens the analysis substantially.

We gained an introduction to the digital humanities through a workshop sponsored by the National Humanities Center, a private U.S. organization in North Carolina. (Ricœur himself spent two academic years at this institution while working on the initial volume of *Time and Narrative*.) We were two of fifteen scholars chosen to participate in the workshop, and we convened for two week-long periods in June 2015 and 2016. We undertook extensive background reading in advance and spent most of the week the first year learning the computer programming language R which is one of the most robust computer languages for text analysis. During our second summer we discussed drafts of our various projects, including our computer experiments.

Our partnership proved to be a valuable asset. While we are both experienced Ricœur researchers, we both were new to the digital humanities. Fernando picked up R programming very quickly. He is not only a professor of philosophy in Brazil, but at the same time also worked for years as a software engineer (and manager) for Motorola, a position he only terminated in May 2016. Programming is a natural language for him. George, by contrast, was new to programming and needed to think through the details of the experiments. Fernando's facility with programming combined with George's perspective as the new user who can ask questions

or seek clarifications have allowed us to begin to generate experiments that we think will be useful as models for other Ricœur researchers who have no background in programming or in R, as we return to in part 5.

We offer as a reminder about our digital humanities experiments that we have concentrated more on developing the search logic than being concerned about their ultimate accuracy. We have not focused on completing the database or reviewing plain texts. We have run the experiments with these limitations in mind in order to give more attention to fine tuning the experiments' parameters.

We describe in narrative form the experiments here and do not include any technical detail about the search logic.

a) Word Count

For this experiment, we used as a sample one of Ricœur's most well known texts, *Oneself as Another*. We created a NLP tagged text, filtered it to retain only common nouns, counted all occurrences, and plotted the use curve.
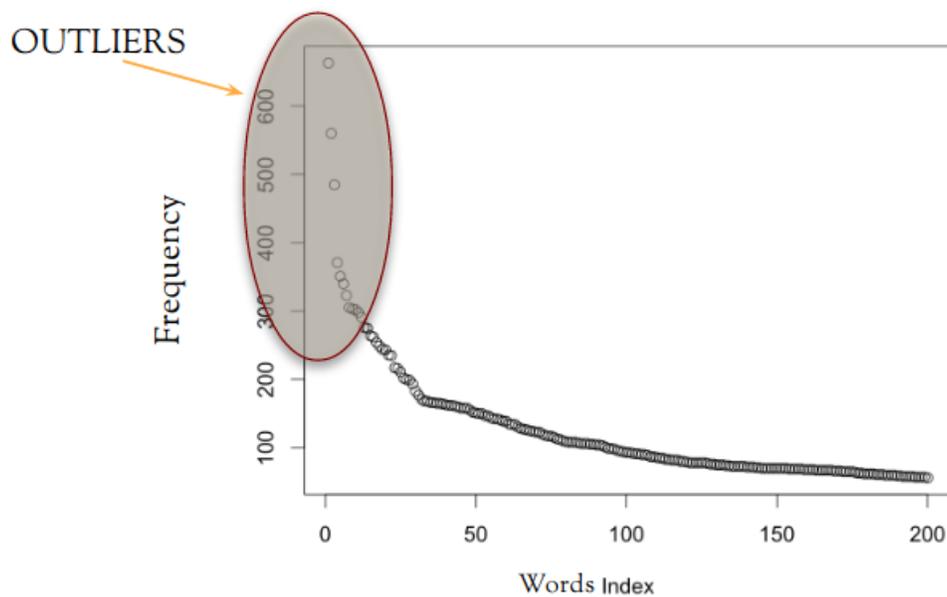
The outlier words have the following number of occurrences:

action	other	sense	self	same	question	way	identity	narrative	first
663	560	485	371	351	340	323	305	303	302
character	relation	person	time	fact	life	justice	moral	theory	notion
298	291	276	275	264	262	254	249	244	243

We were particularly fascinated here to see that in a book on the self, "action" was by far the most prominent word. Other terms whose prevalence struck us were "narrative" and "justice." Both are significant themes in other parts of Ricœur's work, and the volume of references to "narrative" in particular showed a relationship to other of Ricœur's texts that we had not anticipated.

b) Unsupervised Clustering: Dendogram

Cluster dendograms show the proximal relationships between texts and has been used, for example, to determine the likely author of an anonymous text. We initially undertook a cluster dendogram analysis of Ricœur's 50 longest works (to which we had digital access), but the dendogram was too cluttered to reproduce easily, so we reduced the dendogram to the 20 longest works. For this experiment we created a NLP tagged text, again filtered it to retain only common nouns, clustered texts based on the remaining nouns, and generated a dendogram with the results.

Cluster Dendrogram
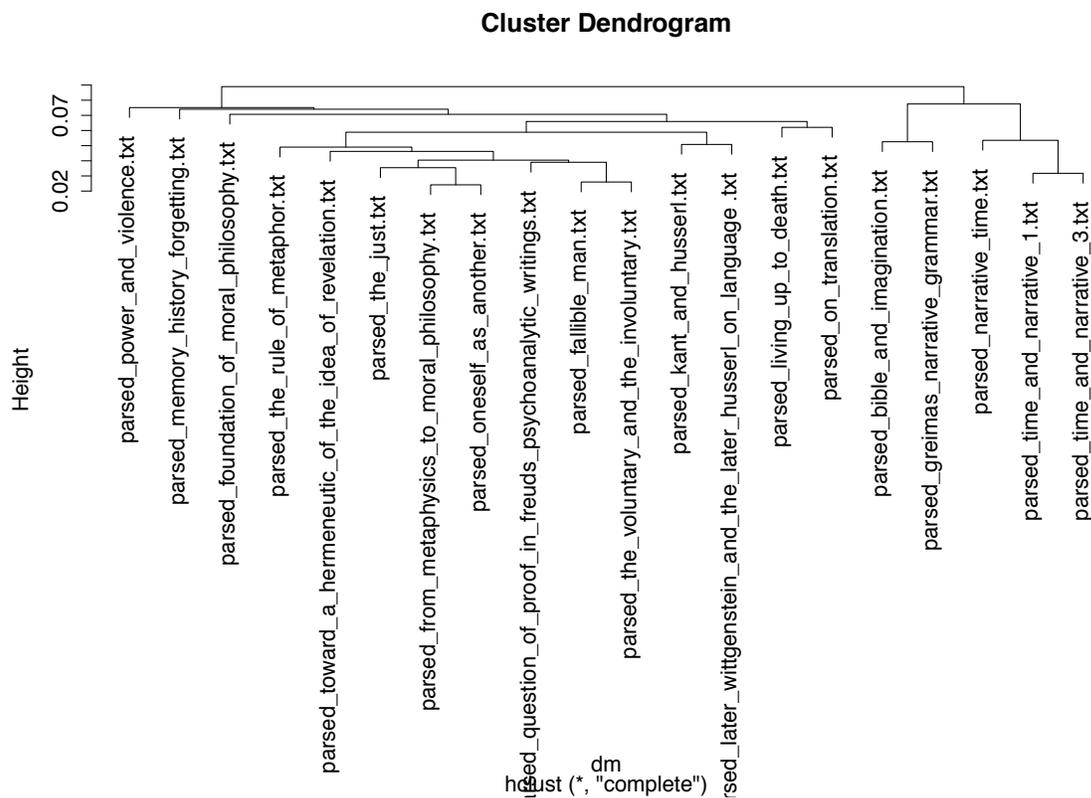
Some of the relationships were unsurprising, such as the relationship between volumes 1 and 3 of *Time and Narrative* (on the far right), but some were unexpected, such as the relationship between *The Rule of Metaphor* and the article on "Toward a Hermeneutic of the Idea of Revelation" (on the near left). This relationship is intriguing, because a major concern of the metaphor book is to show how metaphor allows the creative insight to break into the structure of existing language. Metaphor may be considered a secular form of revelation. Metaphor may retain relevant as a form of inspiration even for those for whom religious revelation no longer seems meaningful. Conversely, the juxtaposition between metaphor and revelation as both creative may help revelation retain salience in the contemporary age. Further, the role of metaphor in religious language raises a host of provocative issues much debated in the literature: is religious revelation the product of the divine or the product of metaphoric language? We have not had time to consider the implications of texts that are not in close relationships.

c) Ricœur's Social Network

This experiment assesses those thinkers with whom Ricœur's work interacts the most over time. We created a NLP tagged text, filtered it to retain only proper nouns, counted the most frequent references to thinkers in the first graph, and in the second graph grouped works by decades and counted the frequency of references to thinkers by decades.

Ricoeur's Social Network - Overall
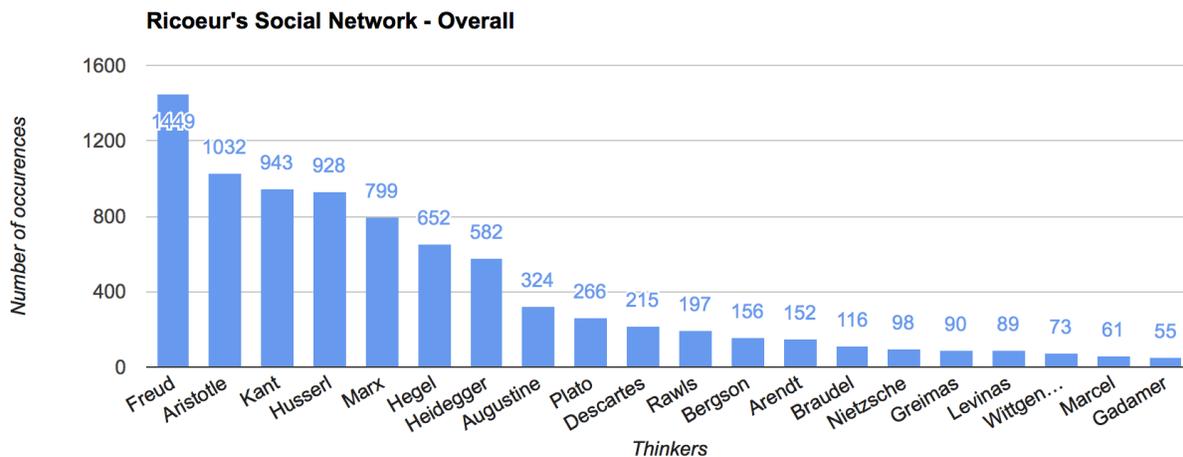
Although the graph is a product of the limited works to which we had digital access, the Network was particularly striking in the lower numbers. We would not have expected so many references to Wittgenstein. We were surprised to see that references to the semiologist Greimas outnumbered those to the hermeneutic scholar Gadamer.

Ricoeur's Social Network Evolution by decades
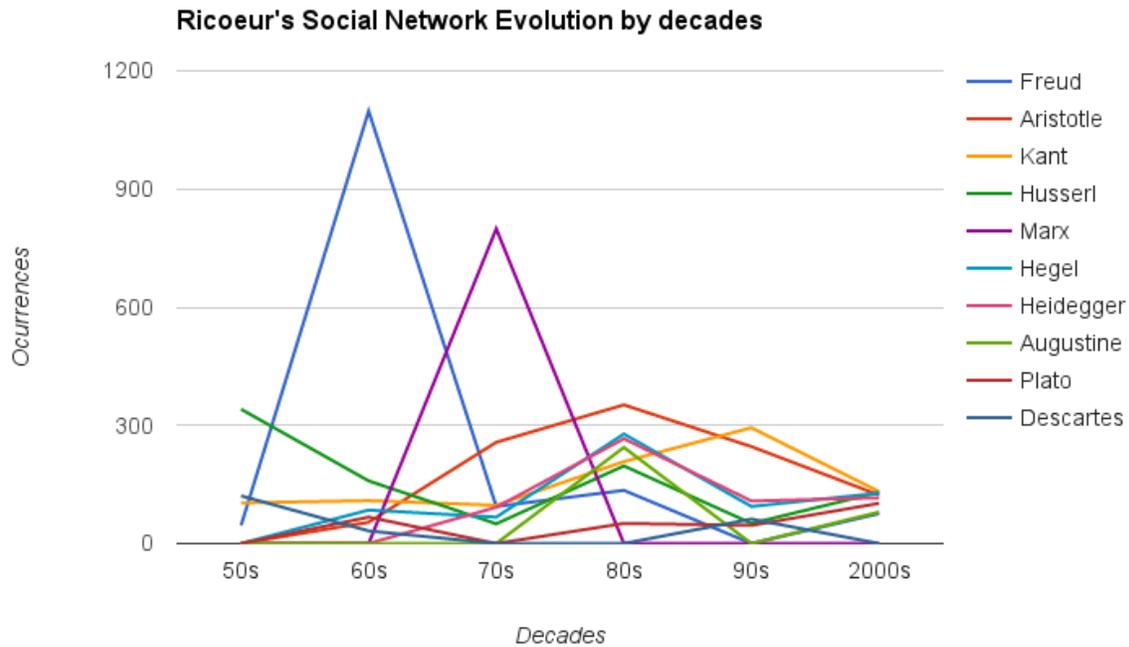
This graph complied in many ways with our expectations as readers of Ricœur. Although largely unsurprising, the experiment was useful to provide empirical confirmation of our more individual impressions. One of the values of digital humanities is that it can act valuably to confirm or disconfirm more subjective orientations. One reference that was unanticipated was the continuous recourse to Husserl. The general expectation is that Ricœur turned away from his phenomenological roots in Husserl in his hermeneutic turn in the 1970's. It was also valuable to analyze this graph in relation to the prior graph on Ricœur's social network. One might think from the prior graph that Freud (listed first) and Marx (listed fifth) were some of the most prominent influences on Ricœur. In the present graph, though, readers of Ricœur will know that the spikes in Ricœur's references to Freud and Marx were largely the result of two books focusing, respectively, on each figure: *Freud and Philosophy* and the *Lectures on Ideology and Utopia*. While over Ricœur's career Freud and Marx remained significant influences, of greater relevance were those figures to whom the graph shows Ricœur gave more constant attention over the decades: Aristotle, Kant, and Hegel.

d) Most Frequent Words over Time

Here we created a NLP tagged text, filtered it to retain only common nouns, counted the most frequent words, grouped works by decades, and counted the frequency of words by decades.

Decade	1 word	frequency	2 word	frequency	3 word	frequency	4 word	frequency	5 word	frequency	6 word	frequency	7 word	frequency	8 word	frequency
50s	consciousness	1114	body	696	freedom	656	sense	619	action	556	object	548	world	548	meaning	537
60s	meaning	733	man	590	language	562	desire	554	reality	496	psychoanalysis	489	ego	483	sense	476
70s	metaphor	1949	meaning	1396	ideology	1370	discourse	1293	language	1224	concept	1109	sense	1047	theory	914
80s	time	4384	narrative	2113	history	2108	way	938	sense	927	action	865	past	855	work	852
90s	action	835	sense	834	justice	789	question	699	time	698	way	662	idea	604	relation	520
2000s	die	4071	Memory	2267	History	1672	time	1239	recognition	810	way	701	work	691	sense	655

The larger issue here is of special interest to Ricœur scholars. Vocabulary he might come to focus on in later years, for example, on "recognition," might be anticipated in earlier studies. Drilling down into this list might be very useful for this kind of inquiry. Of particular interest in the list presented were terms such as the prevalence of attention in the 1950s to the "body" (an element of Ricœur being appropriated by contemporary feminists), and the attention to "language" in the 1960s, in advance of his more general hermeneutic turn to language in the 1970s. As expected, in the 1970s some of the most prevalent vocabulary was "metaphor," "discourse," and "language."

e) Topic Modeling

This experiment allowed us to examine Ricœur's major topics across his corpus and the network of words surrounding each of these topics. Here we created a NLP tagged text, filtered it to retain common nouns, topic modeled the entire set of texts, and created word clouds for all topics. In general it was quite fascinating to see the word cloud themes that the software came up with. The prominence of attention in the word clouds to history, time, justice, language, and identity were not surprising and reflected well Ricœur's corpus.

geschichte retrospection cultures
claims antiquity expectations
weinrich right epoch histories position
volume art
generation judgment traditions
promise consciousness claim enlightenment
hope futures essay philosophy progress ambition
truth times iii expectation end singular
plea historian
ages loss modernity present life crisis gadamer
justice nietzsche idea time future reason tone
realization france
theme
critique liistory belief
age era
excess
course space past hegel nation rupture
ideas
continuity spirit horizon koselleck force article
german culture
narrative experience term reflection realms
habermas categories condition hermeneutics
humanity concept century revolution reception
chronology historiography modern
legitimacy archeology identity novelty
discussion yerushalmi effect renaissance
temporalization reinhart

beingtowarddeath narrated deatli primacy
features possibilities phenomenon emphasis
ecstases science variations geschichtlichkeit
finitude authenticity end resoluteness beingawhole
things possibility repetition division tales
thinking analyses historicality ontology ordinary
historicity past dasein chapter yorck
destiny history temporality problems
face structure
day concept term debt
condition future time world mode
level
concern basis death section
meaning fate
unity life heidegger derivation
heritage analysis care present sciences turn
conception phenomenology theme thrownness
augustine withintimeness existentiell nows
quest tem preoccupation beingtowardsdeath
epistemology understanding anticipation datability
beingintheworld historiography interpretation
husserl dilthey temporalization entities fallenness
horizon reference

freedom
differences positions instances presupposition
societies agreement equilibrium advantages
decision utilitarianism deliberation autonomy
domination procedure convictions version
authority judgment a right institutions p inasmuch
structure conception goods position doctrine
conditions bond division
john good city idea situation shares
distribution theory order violence
equality diversity
fairness interests
conflict justice rights claims
choice ignorance cities
share argument rawls society veil liberty
victim spheres principles state injustice
question rule persons
reasons principle contract parties
institution justification book foundation
democracy individuals walzer everyone
arguments consensus possible
inequalities plurality system competition
constraint argumentation partners individual
habermas nature impartiality

berman meanings
schleiermacher equivalence alternatives
exchange belongingto construction appendix
dialogue relationship translator desire
rhetoric signs languages to community
plurality issue world meaning exegesis
works latin english
task language earth
source thesis
act self essay
verb title work text page bible
book texts thing
love word talk
translation test gap
real ricoeurs
circle interpretation french
ricoeur words sense
translators reader understanding author letter
paradigm phenomenology idealism target
ideologies hermeneutic gadamer heidegger
appropriation explication distanciation diversity
translations distantiation babel dilemma

In the word cloud on history, we noticed in particular the relationship between history and hermeneutics.

In the word cloud on language, of particular interest was the conjunction between language and phenomenology. Often phenomenology is tied to perception rather than to language. Also of interest was the relationship between phenomenology and hermeneutics; hermeneutics did not separate itself from phenomenology, as many would believe.

In the word cloud on identity, prominent was the conjunction between identity, self, and body, which goes beyond the Cartesian duality between mind and body, an important theme in Ricœur.

The word clouds may be of particular value to those new to Ricœur, such as graduate students. The word clouds offer a valuable overview of Ricœur's main themes and then correlate subordinate themes that suggest potential topics worthy of greater scholarly attention.

g) Word Associations in Context: Genesis, Formation, and Changes in Concepts Over Time

We would like to establish searches to assess the genesis, formation, and changes in Ricœur's thought about particular topics over his 60 years as an author. Potential topics include his work on hermeneutics, recognition, justice, history, culture, imagination, and the self. We would run experiments evaluating the most frequent 100 words surrounding each term and then divide the analysis by decades. Thus far, we have analyzed the term "hermeneutic*," which includes hermeneutic, hermeneutics, and hermeneutical. Here we created a NLP tagged text, filtered it to retain only common and proper nouns, plurals and singulars, and adjectives or numerals. The algorithm then counted each of the 25 terms before and after each occurrence of the keyword. It also avoided double counting areas by making sure that a text region already considered around a given keyword occurrence was not considered for another occurrence of the keyword in the same region. Finally, the algorithms summed up all occurrences of each term and sorted the result decreasingly by the total number of occurrences. The following provides an edited sample of the experiment and shows the highest frequency words in each decade surrounding "hermeneutic*":

1950s	ricoeur	phenomenology	hermeneutics	hermeneutic	meaning	narrative	
	27	13	10	7	7	7	
1960s	hermeneutics	interpretation	meaning	reflection	freud	symbols	language
	174	157	115	110	84	73	72
	hermeneutic	psychoanalysis	problem	first	philosophy	consciousness	symbolism
	61	56	54	50	47	41	40
1970s	hermeneutics	text	discourse	language	interpretation	hermeneutical	world
	218	125	111	92	84	67	66

	first	historical	phenomenology	sense	reference	hermeneutical	world
	63	63	62	62	61	67	66
1980s	time	hermeneutics	hermeneutic	Ricoeur	historical	text	meaning
	260	242	183	123	118	115	113
	phenomenology	history	interpretation	testimony	question	understanding	sense
	112	109	104	94	90	88	86
1990s	hermeneutics	other	self	way	action	question	philosophy
	104	54	50	46	43	43	41
	understanding	sense	first	work	explanation	time	theory
	40	35	33	33	32	43	41
2000s	die	history	hermeneutics	**tlie	memory	historical	condition
	107	81	80	66	62	54	31
	ricoeur	critical	phenomenology	time	philosophy	hermeneutic	human
	31	28	28	28	27	24	23

In the 1950s, Ricœur's use of the term was low and was mainly associated with the term "phenomenology," which we expected. At this stage, Ricœur's work was more overtly phenomenological. In the 1960s usage of hermeneutics was much higher and tied into terms such as "interpretation" and "meaning" and also related to "Freud" and "psychoanalysis." During this period Ricœur was developing his hermeneutic approach through an investigation of Freud and Freud's hermeneutic of suspicion. In the 1970s usage revolved around "discourse" and "language" as Ricœur turned to a hermeneutics of language more broadly, and so a broadening beyond the sphere of psychoanalysis. In the 1980s terms such as "time" and the "historical" became prominent. Interestingly, "phenomenology" also resurfaced as a major term; Ricœur did not set aside his phenomenological beginnings.

We would ultimately like to determine how Ricœur's definition of hermeneutics may have developed over time. The word association analysis provides some general assistance in this text, but closer textual analysis would be necessary to explore these changes with precision.

h) Combine Word Citations with Secondary Sources

We have run initial experiments that assess major secondary sources in Ricœur – such as Aristotle, Kant, and Hegel – and analyze the word usage around these references to help detect possible origins of Ricœur's vocabulary and concepts. We initially analyzed the 50 words surrounding each figure in the entire Ricœur corpus available and established lists of the top 100 terms. The algorithm followed the same basic steps described in section 4b. Here, for example, are the top terms in the list for Aristotle:

aristotle	time	metaphor	action	other	narrative	sense
1095	545	403	281	272	265	265
rhetoric	discourse	poetics	theory	first	same	way
231	223	220	201	199	199	195

We did not find this list very informative or surprising. We then narrowed the search to the top 100 terms surrounding principal secondary sources in one representative Ricœur text, *Oneself as Another*. Here are the top terms for that search for Aristotle:

aristotle	action	other	sense	good	life	justice	self
205	92	64	62	56	52	45	42
man	praxis	narrative	ethics	things	relation	way	character
41	41	40	38	38	37	37	35
end	friendship	power	first	choice	agent	term	analysis
35	35	35	32	31	30	30	29

Interesting for us here were that the top terms closely reflect a major theme in this text, that the ethical intention is aimed at the good life with and for others in just institutions. The data show the close ties of this theme to Ricœur's appropriation of Aristotle. We then narrowed the search even further, to the principal chapter on Aristotle in *Oneself as Another*, which led to the following main results:

aristotle	good	other	friendship	life	self	man	sense
58	38	31	30	30	28	27	25
justice	end	ethics	things	action	aim	relation	way
22	20	16	16	15	13	13	13

We compared this list to the top 100 words in Aristotle's *Nichomachean Ethics*, from which Ricœur principally draws. The top terms here were:

man	good	things	such	men	pleasure	other	reason
840	545	379	357	335	288	281	250
same	friendship	kind	selfcontrol	case	way	state	virtue
242	195	182	180	172	167	166	160

We found these results quite revealing. Ricœur's vocabulary demonstrated, for instance, the teleological dimension of his interest in Aristotle in this chapter ("good," "life," "end," "ethics," "aim"). The commonality of the usage of "friendship" confirms other scholarship that highlights the role of this term in both of these thinkers. It was also illuminating to contrast the vocabulary between the two thinkers. The term "mean" was the 17th occurrence in Aristotle but did not make the top 100 in Ricœur. Ricœur may drop reference to the "mean" because he replaces this criterion with Kant's formal test of the golden rule (the subject of the next chapter in Ricœur's text). It was also interesting to us that in Aristotle's vocabulary list the notion of evil does not rise prominently. Instead there are terms such as "bad" and "pain." While for Ricœur, the tragedy of human action is a continued focus (and is a primary subject of a succeeding chapter), Aristotle's ethics is known for a more optimistic approach to anthropological goodness and self-realization.[8] This juxtaposition between Ricœur and Aristotle is very helpful in again suggesting why Ricœur turns to Kant to mediate Aristotle. This topic has drawn significant attention in Ricœur scholarship, and the more delineated comparison of vocabulary renders the juxtaposition more precise. It was also interesting to us that Aristotle uses more the vocabulary of "self-control" and "self-mastery," while Ricœur does not. We would want to consider this difference at greater length, but it may that Ricœur's attention to the tragic, to the psychoanalytic, and to the diversity of the self ("oneself as another") renders self-mastery too utopian a goal.

We intend to run a similar experiment in *Oneself as Another* to establish comparisons between the chapter on Kant and Kant's vocabulary in the *Groundwork of the Metaphysics of Morals*.

More generally, it was significant to us that the comparisons between Aristotle and Ricœur that we were able to draw were due to our having already considerable familiarity with the texts in question. A newcomer to this material would likely not be as aware of the relevance of similarities or divergences in the vocabulary between the two thinkers.

i) Distinguish Positive and Negative References to Ricœur's Social Network

Our earlier experiment generating evidence of Ricœur's principal interlocutors in his work did not clarify the nature of his response to these thinkers. We decided to investigate sentiment analysis software to determine whether it would help differentiate between Ricœur's positive or negative reference to a secondary source. We ran experiments again using *Oneself as Another* and ran individual experiments focusing on Ricœur's references to Aristotle, Kant, and Hegel. Here we used the software to parse the text and extract sentences, filtered the sentences that contained the thinker's name, used the software algorithm to extract the sentiment weights for each sentence, and calculated the mean and average for the vector. We initially used the Syuzhet sentiment analysis package, but found that the assessments did not distinguish between Ricœur's evaluation of the thinker and his presentation of the thinker's own views. We then tried another sentiment analysis package, SentiStrength, and we did find that software more useful. SentiStrength evaluates the strength of both positive and negative sentiments and provides a score on each measure for every sentence. Scores run from -1 (not negative) to -5 (extremely negative) and from 1 (not positive) to 5 (extremely positive).

Here are two representative samples from the analysis of Hegel and then one from the analysis of Aristotle. In each the software is analyzing one sentence of Ricœur's text.

[16] "One thing should be said straightaway: if one must, at some point, \"renounce Hegel,\" his treatment of tragedy is not the place to do so; if Hegel is reproached with having imposed a \"synthesis\" upon all the divisions that his philosophy displays, the genius for discovering or inventing this synthesis is certainly not drawn from tragedy."

1   -3   [16][0] One[0] thing[0] should[0] be[0] said[0] straightaway[0] if[0] one[0] must[0] at[0] some[0] point[0] \[0] renounce[-1] Hegel[0] ,\[0] his[0] treatment[0] of[0] tragedy[-2] is[0] not[0] the[0] place[0] to[0] do[0] so[0] if[0] Hegel[0] is[0] reproached[-1] with[0] having[0] imposed[-1] a[0] \[0] synthesis\[0] upon[0] all[0] the[0] divisions[0] that[0] his[0] philosophy[0] displays[0] the[0] genius[0] for[0] discovering[0] or[0] inventing[0] this[0] synthesis[0] is[0] certainly[0] not[0] drawn[0] from[0] tragedy[-2] [[Sentence=-3,1=word max, 1-5]][[[1,-3 max of sentences]]]

The analysis properly scores Ricœur's critique of Hegel ("renounce"), yet marks as negative the references to tragedy. This seems too simple an evaluation. Tragic action is of course in one sense negative, yet in a story – Ricœur's example is Hegel's analysis of the tragedy of Antigone – may have positive lessons. Also, Ricœur is applauding Hegel's treatment of tragedy as something unsurpassable. The software does not distinguish between Ricœur's evaluation of Hegel and the presentation of Hegel's own views.

[50] "The artifice of the Hegelian construction is not, however, to be deplored; as an artifice it takes its place among the excesses, transgressions, and hyperboles of all sorts that nourish moral reflection and, perhaps, philosophical reflection in general."

1   -2   [50][0] The[0] artifice[0] of[0] the[0] Hegelian[0] construction[0] is[0] not[0] however[0] to[0] be[0] deplored[-1] as[0] an[0] artifice[0] it[0] takes[0] its[0] place[0] among[0] the[0] excesses[0] transgressions[-1] and[0] hyperboles[0] of[0] all[0] sorts[0] that[0] nourish[0] moral[0] reflection[0] and[0] perhaps[0] philosophical[0] reflection[0] in[0] general[0] [[Sentence=-2,1=word max, 1-5]][[[1,-2 max of sentences]]]

It was odd to us here that the software did not mark "artifice" as something negative, even though Ricœur's assessment is more positive. The software also marked "deplored" as negative, even though it is preceded by "not." Again the software could not apparently read Ricœur's treatment as a positive assessment.

[10] "Nevertheless, in order to have a point of anchorage on the level of action for [Aristotle's] detailed study of virtue, that is, of the features of excellence belonging to action, he undertakes in book 3 of the Nichomachean Ethics an initial distinction pairing actions performed, despite oneself . . . with those performed freely . . . ."

4   -1   [10][0] Nevertheless[0] in[0] order[0] to[0] have[0] a[0] point[0] of[0] anchorage[0] on[0] the[0] level[0] of[0] action[0] for[0] [Aristotle's][0] detailed[0] study[0] of[0] virtue[2] that[0] is[0] of[0] the[0] features[0] of[0] excellence[3] belonging[0] to[0] action[0] he[0] undertakes[0] in[0] book[0] 3[0] of[0] the[0] Nichomachean[0] Ethics[0] an[0] initial[0] distinction[0] pairing[0] actions[0] performed[0] despite[0] oneself[0] . . . with[0] those[0] performed[0] freely[0] [[Sentence=-1,1=word max, 1-5]] [[[4,-1 max of sentences]]]

The grading of Ricœur's assessment of Aristotle is very positive, yet again Ricœur here is not presenting his own evaluation of Aristotle but describing the nature of Aristotle's own ethical approach, which is focused on virtues.

We should acknowledge, as its website indicates, that SentiStrength was designed not to analyze longer or more complicated texts but rather to evaluate the emotion of authors toward a topic being discussed in short texts such as social media (e.g., Twitter feeds or Facebook postings). For us, the sentiment analysis did help locate statements in the text presenting stronger sentiments, but the analysis was not very deep. At this point, we have not found sentiment analysis to be that productive at least for assessment of philosophical material. Again the larger issue here is the use of a digital humanities tool to differentiate between an author's positive or critical portrayal of another thinker. It may be that we need to conjoin with sentiment analysis other software tools that offer more sophisticated treatment of grammar and syntax.

Conclusion to Part 4

The experiments we have run thus far have offered much food for reflection. A number of the experiments we have conducted have yielded significant results that are quite insightful in helping us better to understand the nature and import of Ricœur's work. Some of the more delineated experiments seem particularly of value for more established Ricœur researchers. Researchers newer to Ricœur would benefit from the topic analysis and the general social network analysis. We are grateful for suggestions of experiments provided us both by our digital humanities teachers and by audience members at our conference presentations. We greatly appreciate the audiences' participation.[9] Over time we hope to consider more what other kinds of experiments might be distinctively pertinent to philosophical texts.

We remain attentive to the need to keep in mind the distinction between the complexity of the digital humanities analysis and the substantiality of its results. Interestingly for us, one theme of Ricœur's work is helpful with this task, his distinction between explanation and understanding.[10] We want to explore the value to the digital scholar of the stage of explanation – the use of the tool – but the tool is of value not in itself but only so long as it assists our understanding, assists our insight into what it means to be human. Digital analysis provides a vast amount of data, but it is left to us to analyze the significance of this data.

5. Archive Analysis Portal

As a final stage of the larger digitization project, we are working on creating a web portal interface for the digital Ricœur archive that would allow Ricœur scholars to enter variables into pre-established search logics, so that they can receive the benefits of digital humanities searches without having to learn the programming themselves. Along the lines of the experiments discussed in part 4, we would create the search logic programs, the portal design would allow researcher input, and then the search logic would operate in the background behind the portal. The portal will be deployed in a three-tier architecture commonly known as MVC (model, view, controller). The *model* layer contains the system's database, and for us this will include the files with Ricœur's works. These files will be initially of two different types, raw text files and Stanford NPL tagged files. The *controller* layer implements the algorithms that process the data in

order to provide useful information to the end user. The *view* layer presents the data to the user. We would of course be open to modification and expansion of the search logic programs as researcher interest indicated. This portal interface and its programming may be a model for use not only in Ricœur studies but in other philosophical or other digital scholarly research programs.

For the development of the portal, we have added to our team Ming Yeung Cheung. Ming Yeung is a native of Hong Kong who is a Ricœur scholar now teaching religious studies at Fu-Jen University in Taiwan. Before he switched to graduate training in theology, he had graduate training in and several years experience working as an electronic engineer.

The portal team has been exploring the option of using Voyant or Shiny software for the web interface. We are beginning with use of Voyant and then over time will consider migrating to Shiny as the complexity of demands on the portal increases, since Shiny permits more programming flexibility. Shiny helps make R code more dynamic and interactive, but being a general purpose web interface for R programming, it requires some effort for deploying text-based experiments via customized widgets. We may not be able to do with Shiny everything we want (due to the limitations inherent in any software tool). By contrast, since Voyant is specifically designed for text analysis, it offers a smoother entry path for deploying simple pre-defined text analysis widgets, such as word clouds and word counting. Once we determine with Voyant which widgets are most useful, we then may move on to Shiny. To provide copyright protection to the original full texts in the database, we will establish portal access control policies that will prevent a user from reconstructing a work by systematic searches.

The digital Ricœur portal will be the main Internet interface for accessing the analysis tools exposed by the system. It will be hosted on a secure web server along with the other software components entities. The portal's main functionality is to provide an easily accessible façade for the tools that we will make available to the scholarly community. It also provides a level of flexibility for future enhancement and integration of widgets other than the initial planned ones in Voyant and Shiny.

Scholarly use of the web portal to run searches may itself provide additional benefits to the Ricœur scholar community. For example Ming Yeung has suggested that, with the search author's permission, we might be able to tag searches undertaken so that these might be identified for other researcher use. Further, if the author wanted, he or she could opt to identify themselves so that others could contact them for additional scholarly dialogue or collaboration.

Conclusion

Our digitization project and its digital humanities focus seem to have potentially large ramifications for Ricœur studies and for the continuing growth of those studies. The digital access to the Ricœur archive and to the tools to explore it should facilitate and enlarge the depth and scope of Ricœur scholarship. The project also has the important potential to further a sense of international scholarly community and collaboration among Ricœur scholars. The potential for community and collaboration are no small feats in a world where Ricœur scholars are scattered in universities across the globe and, more ominously, the humanities, philosophy, and hermeneutics

are all under threat due to internal and external pressures for universities to direct students into allegedly more economically productive disciplines.

For anyone interested in participating in the elaboration or testing of this digital project, please email co-author Fernando Nascimento (fern.nascimento@gmail.com).

[1] Frans D. Vansina, ed. *Paul Ricœur Bibliographie Primaire et Secondaire/Primary and Secondary Bibliography 1935-2008*, (Leuven: Peeters, 2008). Since 2008, the Fonds Ricœur has been providing more recent bibliographic updates at its website online (www.fondsRicœur.fr).

[2] Ricœur's literary executors hold the copyright for most of Ricœur's publications. Formally these executors are known as the *Comité éditorial* of the Fonds Ricœur, but for brevity we shall refer simply to the Fonds Ricœur. To the extent that the Fonds Ricœur does not hold the copyright – for example, in some translations of Ricœur's work – we would still want to reach agreement with the Fonds Ricœur about their digitization as a matter of comity.

The Fonds Ricoeur has been engaged in a multi-year project that will extend for approximately another two years to digitize materials by Ricoeur in its own archive. Under an initial grant from the French government, the digitization was undertaken primarily as a matter of security, to avoid the loss of material in case of a calamity such as a fire. Under a new grant from l'Université Paris Sciences et Lettres, digitization will continue and make available to the public online approximately 300 texts, largely Ricoeur articles not collected in published volumes.

[3] Authors Guild v. Google, 804 F.3d 202, 229 (2d Cir. 2015).

[4] Authors Guild v. HathiTrust, 755 F.3d 87 (2d Cir. 2014). The *HathiTrust* decision was not appealed for review by the U.S. Supreme Court and so remains good law.

[5] Authors Guild v. Google, 804 F.3d 202, 216-18, 221-22, 223, 224-25 (2d Cir. 2015).

[6] *Ibid.* at 223. The losing authors in the *Google* appellate case subsequently appealed the decision to the U.S. Supreme Court, and the Supreme Court declined to review the appellate court judgment, thus ending the case. *Authors Guild v. Google*, 136 S.Ct. 1658 (April 18, 2016). See also Adam Liptak, "Challenge to Google Books is Decline by Supreme Court," *New York Times*, April 18, 2016. Review by the Supreme Court in these kinds of cases is discretionary. The Court's decision not to grant review should not, however, be interpreted to imply the Court's agreement with the appellate court's decision. The Court typically, as here, does not indicate its reasons for denying review, and many potential reasons for the refusal exist, including the Court's desire to wait for further case developments in the area. *See, e.g.*, *Teague v. Lane*, 489 U.S. 288, 296 (1989) (observing that the Supreme Court's denial of an appeal "'imports no expression of opinions upon the merits of the case.' The 'variety of considerations [that] underlie denials of the writ' counsels against according denials of [appeals] any precedential value") (citations omitted). In any event, the Second Circuit decision currently remains good law. In the future, the rationale for the decision could be challenged by plaintiffs in other cases in other circuits of appeals, and the Supreme Court could decide to review a subsequent case, which would bring the Second Circuit decision back into question.

[7] Franco Moretti, *Distant Reading* (London: Verso, 2013).

[8] In Ricœur's chapter in *Oneself as Another* that is the basis for the present experiment, the only text of Aristotle that Ricœur draws upon is the *Nichomachean Ethics*. It therefore seems appropriate to limit the comparison between Ricœur and Aristotle to their ethics. The subject would be more complicated

if we were to expand attention to Aristotle's discussion of the tragedy in his *Poetics*, a text that Ricœur also addresses elsewhere.

9 We are especially grateful for the engagement of a counterpart from the Fonds Ricœur, Olivier Abel, who was a speaker at one of the conferences at which we also presented and was learning through our presentation there of the details of the digitization project for the first time. He offered several useful suggestions for experiments himself. One proposal that Olivier raised that we have not had the chance to pursue is to distinguish between citations arising from Ricœur's major texts and those from his minor ones.

10 See Paul Ricœur, "What is a Text? Explanation and Understanding," in *From Text to Action*, Kathleen Blamey & John B. Thompson trans., (Evanston, IL: Northwestern University Press, 1991), 105-24; "Explanation and Understanding: On Some Remarkable Connections Between the Theory of Texts, Action Theory, and the Theory of History," in *From Text to Action*, 12-43.