Type of report:	Final performance report
Grant number:	HD-51128-10
Title of the project:	Mapping the History of Knowledge: Text-Based Tools and Algorithms for Tracking the Development of Concepts
Project directors:	Mikhail Gronas, Anna Rumshisky
Grant type:	Digital Humanities Start-Up Grant (Level II)
Institutions:	Dartmouth College, Brandeis University
Date submitted:	January 21, 2013

1 Introduction

This report describes the tasks completed for the first stage of the project that investigates the possibility of using historical digitized text to track long-range historical change, and specifically, the history of intellectual networks and reputations. The overall aim of the project is to map out the history of representation of knowledge in Europe over last three centuries using as a proxy the history of changes in historical editions of Encyclopedia Britannica. In this first stage, we focused specifically on the tools for tracking and visualization of the relative importance of people, interconnections between them, and the rise and fall of their reputations. Over the course of the project period, we performed a series of corpus-analytical tasks that were necessary for building the analytical and comparative tools for historical analysis using scanned noisy text of historical editions.

2 Project Activities

We used the corpus of several historical editions of encyclopedia Britannica; in addition, we used the current edition Britannica and Wikipedia to supplement the analyses. Normalization across editions for all concepts covered in the Encyclopedia proved to be a challenging task due to the noise in the scanned text. We therefore decided to limit the first stage of the project handling the articles about people, thus focusing on the social dimension of the history of knowledge (i.e. the history of reputations and intellectual networks).

Historical editions in our corpus are OCR scans, and therefore contain very noisy data. In order to use these texts for comparative analysis of cross-edition changes, we had to perform a series of corpus-analytical tasks, including:

1. Splitting the text into articles and identifying article titles.
2. Identifying articles about people.
3. Matching the articles across editions.
4. identifying explicit references to other articles.
5. Identifying article categories and matching them across editions.

2.1 Data set

We used three scanned and OCR'ed out-of-copyright editions of Britannica: Editions 3 and 9 by GoogleBooks, and Edition 11 available from jrank.org. Encyclopedia Britannica granted us research rights to use the electronic text of Britannica's current (15th) edition in XML format. We also used Wikipedia, which effectively provides an (extensive) update to the 15th edition, since the original articles from Britannica's older edition often served as the initial version for Wikipedia entries. These editions gave us the total of 5 points of comparison:

1. Edition 3 (1788–1797)
2. Edition 9 (1875–1889)
3. Edition 11 (1910–1911)
4. Edition 15 (Current EB; 1985–Present)
5. Wikipedia (Present)

Our choice of the specific historical editions was motivated by the fact that some of they represent distinct changes in the state of Encyclopedia Britannica. Edition 3 represents the initial period

in the history of Encyclopedia Britannica when it was just being established as an authoritative source. Edition 9, hailed as a Scholar's edition, represented a considerable reworking of the previous editions, with multiple respected authorities in different fields of knowledge contributing articles. Edition 11 again represented a change in the state of knowledge; it was a complete reworking of the Encyclopedia, and remained an authoritative source for several decades.

2.2 Restricting Subject Set

Ideally, the comparative analysis should be conducted using all the subjects for which articles are present in the Encyclopedia, i.e. concepts corresponding to the article titles. This set of concepts should be correlated with and supplemented by the concepts extracted from the text of the articles.

In the present work, we restricted our data to articles on people only for the following reasons. Our analysis is based upon matching of the articles between editions (locating articles on the same topic across editions). Also, in order to conduct the full analysis of conceptual relations and relative importance of different concepts, we detect mentions of different subjects from within other articles, effectively creating a hypertext structure. At present stage, this is exceedingly complicated when dealing with most concepts expressed by common nouns, because of (a) differences in taxonomies: same concept can be part of an article in an earlier edition and has its own article in a later edition b) polysemy: same word can serve as a head word of an article (e.g. nature) and be used in a different meaning (e.g. in proposition by nature). A sturdy disambiguation in such cases is highly problematic. However, proper names (e.g. persons), while still constituting a hugely important domain of knowledge, are less affected by these limitations because : a) persons are usually classified as such; b) namesakes are relatively easy to disambiguate.

2.3 Cleaning up the data

Historical editions in our corpus are OCR scans, and therefore contain very noisy data. A large proportion of all words are mis-scanned, with text segments from different articles interspersed. The initial task was therefore to (1) split each of the historical editions into separate articles and (2) identify titles for each article. Edition 11 was obtained from the jrank.org website, where it was pre-split into articles. Despite being collectively edited it does contain a significant amount of errors. Edition 11 is also in progress of being manually corrected as part of Project Gutenberg. We replaced the first 13 volumes of text with the manually corrected volumes.

For Editions 3 and 9, we opted not to build a classifier for this auxiliary task. Rather, splitting the text into articles and title identification was performed using a set of simple formatting heuristics, such as looking for uppercase strings at the beginning of paragraphs preceded by blank lines; eliminating mis-scanned tables by identifying text segments with comparatively small average line length, etc. This was complemented by an *alphabetic ordering check* on title candidates. The latter entails checking the alphabetic ordering of the set of proposed titles to remove some of the false positives that break the ordering.

Checking that the next article is in the correct position alphabetically is not sufficient, since one mis-scanned title could cause all subsequent articles to be marked as bad. For example, if there were three articles in a row titled "AARDVARK", "ABLE", and "ACE", they would be all be marked correct because they are in alphabetical order. If, however, "ABLE" was mis-scanned as "AELE" it would still be correct because it still comes after "AARDVARK", but then "ACE" would be incorrect because it should come before "AELE". To remedy this, we first find all possible articles by just searching for upper case letters, then assign each article a score based on how many articles before it are in fact alphabetically before, and how many articles after it are in fact alphabetically after, using a threshold to filter out false positives.

Under this setup, a large group of bad titles could cause many nearby articles to get a low score. We therefore first run title extraction a low threshold to get rid most large groups of bad titles, followed by several runs with higher cutoffs to weed out the remaining stragglers. Alphabetic ordering check also had to take into account miscellaneous issues such as the fact in that in some of the older editions letters U and V, as well as J and I, were used interchangeably.

We conducted some accuracy testing by manually checking the accuracy of the split-and-extraction algorithm on a subset of the extracted articles. The following estimates for error rates were obtained:

Edition 3 (GoogleBooks) 19.0% error rate
Edition 9 (GoogleBooks) 10.1% error rate
Edition 11 (Jrank) 14.7% error rate

The OCR'ed editions are quite noisy, and we conducted some quantitative investigations of correctness with a modified spell-checker tool which relies on the current edition of Britannica as well as on Wikipedia for lexical information. For the editions obtained from Google Books, considering only the tokens consisting of alphabetic characters with punctuation, the percentage of misspelled words varied across volumes as follows:

- 7.1–9.6 % in Edition 3
- 5.3-7.9% in Edition 9

2.4 Processing graph structure from individual editions

We have investigated several approaches to constructing article graphs. For each edition, two main types of graphs are currently constructed:

1. Distance graphs, using co-occurrence statistics
2. Explicit reference graphs

We have developed software that allows experimentation with different weighting techniques to tune ranking and clustering methods, with preliminary results available for PageRank and Markov Clustering on both types of graphs for each edition. We ran PageRank on both types of graphs, producing importance ranks for individual articles within each edition.

For each edition, we also ran Markov Clustering on the explicit reference graph in order to partition the articles into distinct clusters. Inflation rate, a factor affecting the segmentation of clusters, was an important part of the the algorithm. This parameter was determined by observing the distance between two different clusterings (the number of node changes required to convert one clustering into another) and cluster tightness. Cluster tightness was determined by the product of the Jaccard similarity to each article's wikipedia categories and the cluster size. By averaging this score over all clusters for each inflation value, we could objectively select appropriate Markov Clustering parameters for every edition.

The purpose of clustering all articles from every edition is twofold. First, it can serve as a comparison method between articles from within an edition. Second, and more importantly, clusters can track when certain ideas are no longer associated, at least algorithmically, with what it was associated with before.

2.5 Normalization of articles across different editions

We have done cross-edition normalization using Wikipedia categories and the metadata from the current Britannica editions. The cross-edition mapping approach we have been investigating involves the mapping of different categories from the Wikipedia and Current EB to article sets across historical editions. This involves normalization and mapping of article titles to enables the category mapping. Current approaches we are taking involve distributional ranking of article similarity with differential weighting of different article segments, as well as incorporating weighted use of Wikipedia suggestions and targeted Bing searches. We give more detail on this task in the following section.

Cross-edition article matching

We used TF*IDF to obtain weighted word vector representations of each article. Since the beginning of the article, i.e. the title and the introduction, usually contain a concise version of the most important information laid out in the rest of the article we over-weighed the beginning of each article, in particular giving more weight to strong identificators such as personal names, dates, names of the professions.

For each edition obtained, every article is first matched to the Wikipedia article on the same topic. The articles matching the same Wikipedia article are then matched across editions as follows.

Step 1. Finding candidates for matching

First we use a list of all article titles in Wikipedia which is sorted alphabetically. An insertion index is obtained for the spot where the title of the article in question can be inserted while preserving the sorted order of the list. Then the surrounding k articles around the insertion index are added to our candidate list (we used $k = 6$ in the experiments below). We then query Wikipedia and Bing for each title of an article and add the top 5 results of each query to our candidate list. The final candidate list of Wikipedia articles is compiled by resolving redirects, removing missing articles and adding candidates from disambiguation pages.

Step 2. Candidate Comparison

Each candidate Wikipedia article is compared to the article we are trying to match using a cosine similarity measure computed for the corresponding weighted TF*IDF word vectors.

Step 3. Detecting articles about people

We apply Wikipedia's categories to filter out non-person articles. Wikipedia articles about people are often assigned categories specifying birth or death year (e.g. the article about Johann Sebastian Bach belongs to the categories 1685 births and 1750 deaths). If an article is not assigned a birth-year or death-year category, the original article from Encyclopedia Britannica is filtered out.

Parameter tuning

The above algorithm uses the following parameters:

1. number of words overweighed in the beginning of article ($first\_word\_lim$);
2. number of words in the beginning considered to the title, usu. names and years ($title\_word\_lim$);
3. weight factor for the title words ($title\_word\_ow$);

4. weight factor for for years found in the beginning of the article (*year_ow*;
5. weight factor for professions/occupations (e.g. author, poet, tsar) found in the beginning of the article (*occ_ow*);
6. whether a given word would only be overweighed once; e.g. an occupation might be mentioned multiple times (*only_once_ow*)
7. number of words from the article to query Bing and Wikipedia with, along with the title (*num_words_q*)

We tuned the parameters sing a manually annotated set of 100 articles for each edition. A randomly selected set of articles from Britannica obtained from the initial run of the algorithm was manually matched against the Wikipedia articles and used for parameter tuning. The error rate was then estimated on another 100 of manually matched articles. Table 2.5 summarizes optimal parameter settings for Edition 9.

year_ow	*occ_ow*	*title_words_ow*	*once_only_ow*	*first_words_lim*	*title_word_limit*	*num_words_q*
12	12	8	True	150	4	5

Table 1: Parameter values for Edition 9

The estimates we obtained suggest 75% accuracy, with the error rates for the parameters specified above as follows:

1. incorrect matches - 14.61%
2. non-person articles - 5.26%
3. person articles filtered out - 5.26%

2.6 Assigning categories across editions

We used Wikipedia categories to generalize across editions, the rationale is that such a categorization will allow us to track the development of topics, as well as specific articles.

Wikipedia's categories, while benefiting from the wisdom of the crowds, also inherit problems associated with it. A lot of categories are ad-hoc and not every article in Wikipedia has been assigned all categories that it should conceptually have. An alternative is to use Encyclopedia Britannica's internal tagging system used in the current electronic edition. Each article in our corpus is matched to its edition 15 counterpart (using our Wikipedia matching as crutch) and get its categories.

2.7 Alternative matching strategies

In the final months of the project, we continued working to improve the quality of matching across editions, and developed alternative matching strategies, which use metadata in Britannica Edition 15 and the hidden Persondata template in Wikipedia to detect articles about people. Initial noisy match sets are retrieved by using historical edition frequencies of names specified in metadata.

For each candidate pair in the retrieved match set, spelling correction is applied, and high-impact lexical items that the pair has in common are identified by using a combination of TF*IDF scores, normalized by square root of the position of first occurrence of the word in the article. The resulting set of high-impact lexical is then used to assess distance to other articles in the historical

editions, and the best matching article is added to the match set. Actual matches are then selected by checking how often two candidate articles are paired in the resulting match sets.

We are currently working to evaluate and refine this methodology, which seems to produce promising improvements to the original matching algorithm.

3 Accomplishments

he project was intended as a proof of concept for a method of tracking and mapping the history of knowledge that relies an NLP analysis of historical editions of an encyclopedia. We have demonstrated the validity of our main theoretical assumption: the successive historical editions of an encyclopaedia may serve as a proxy for the history of representation of knowledge. Our goal was to use this theoretical approach to develop analytical methods and produce pilot software tools that track and map this historical dynamics within and across domains of knowledge. On a preliminary (pilot) level, this goal has been achieved, in the form of the set of such methods and tools we describe below.

The main deviation from our initial objectives consists in restricting the object of our analysis. We initially planned to apply our methodology to all concepts expressed as encyclopedia entries; eventually, we limited ourselves to th analysis of entries about people, and thus to the social dimension of knowledge, the history of reputations and intellectual networks. This change has been necessitated by the exceedingly noisy nature of the general dataset (see the subsection on dataset restriction above). The quality of the scans proved to be a lot poorer than anticipated based on the current standards of state-of-the-art scanning technologies. Dealing with such an extent of data contamination due to scanning errors was outside of the scope of the project, and the resources dedicated to this problem in the project budget proved insufficient; internal institutional resources were used to involve student programmers to work on this problem.

Our methodology (consisting in matching the articles on similar topics across editions and determining the domains of knowledge by applying Wikipedia categories retrospectively to the on historical Encyclopedia has proved to be adequate for our analysis. We have developed specific NLP based heuristics and algorithms to deal with the challenges presented by these tasks, as described above. Facing the problem of the exceedingly noisy scans, our team has developed an innovative probabilistic spellchecker that targets large size collections prone to multiple OCR errors. The main achievement of the project is the proof of concept website http://www.newmediacenter.ru/knowevo/gravebook/ and the set of tools for analysis and visualization of the history of reputations and intellectual networks (cf. Section 6).

4 Evaluation

We conducted internal evaluation of accuracy of the tools developed for each of the tasks described above. Please refer to the sections 2.3 and 2.5 above.

5 Continuation of the Project

Our project has generated a considerable interest among scholars working in the fields of intellectual history, history of ideas, and in various historical disciplines. This success suggest that the project is well worth maintaining and developing beyond the start-up stage. The project has inspired an international collaboration with two Russian scholarly institutions: The Department of Computational Linguistics at the Higher School of Economics (Moscow), and The Center for

the Study of new Media (NES, Moscow). The former has invited the PI and Co-PI of the project to use the project as an extended case study in the Course on Digital Humanities, the first such course ever offered in Russia; the latter has suggested a collaborative development of the project and is currently hosting the project's mirror website.

We plan to continue developing the following aspects of the project:

- One important source of noise in the current version of our software stems from the rather inexact matching of the same articles across editions. To improve the results, the alternative (more effective) matching strategies developed towards the end of the grant period will be applied to the historical editions of Britannica and the exploratory interface updated with a more accurate match set.
- The pilot interface is only based on 3 historical editions (plus two contemporary Encyclopedias, Britannica and Wikipedia). The sparsity of data points has often led to insufficient or uninteresting results. We are planning to add all available editions of Britannica to our dataset, which should produce a much richer representation of the historical changes.
- As mentioned above, in order to concentrate on the least problematic data, we limited our testing set to the knowledge related to people. We are planning to add selected groups of other key terms, such as geographical locations, names of disciplines and sciences, names of tools, etc. Our hope is that adding these groups to the exploratory tools would provide useful analysis instruments for the history of science, historical geography, and the history of technology.

6 Grant Products

6.1 Software Tools

6.1.1 Knowevo and Gravebook: Browsing interface and query tools for historical reputation graphs and social mapping of the past

One of the results of our project is a set educational online tools for tracking and mapping the social dimension of the history of knowledge about people, or put simply, a history of reputations. The prototype of the online tools are currently transferred to a new domain, with a mirror hosted by the New Media Center in Moscow, Russia, at `http://www.newmediacenter.ru/knowevo/gravebook/`.

The back-end of these tools is a database that contains all articles about people in different Encyclopedia editions (Britannica 3, 9, 11, 15, and Wikipedia). Articles on the same person are matched across editions to compile a master list of people. Each subject is characterized by measures of importance and centrality in respective editions, by the network of co-occurring subjects ("neighbors"), and by the list of categories accompanying this subject in Wikipedia.

(1) Knowevo reputation graph

This tool maps historical changes in individual reputations. The user inputs the name of the person of interest, and then, if needed, is offered a choice among namesakes. Once the name is selected, the tool produces a graph that represents the importance of this person in successive historical editions of Britannica and in Wikipedia.

(2) Knowevo domain graph

The user picks the domain of interest. The domain is effectively, a Wikipedia category, or a cross-section of Wikipedia categories or lists (e.g. French 19th century composers, chemists, members of

the romantic movement, etc). Once the domain is picked, the system produces the historical graph of the changes in relative importance of the participants of the domain through time, across the historical editions of the Britannica and for Wikipedia. The resulting cumulative graph represents the history of the domain, i.e the historical changes in composition and relative hierarchies within the domains.

(3) Gravebook: Social mapping of the past

One of the products of our research is a novel education tool that highlights and makes more accessible for students the social and intellectual networks of the past. This tool allows one to investigate social and intellectual connections of persons featured in the Current Edition of Britannica and Wikipedia, and reconstruct the underlying social graph, thus creating, effectively, a facebook for the past, or as we term it, Gravebook, an entertaining interface for studying history of human connections. The interface is built as follows. We first determine all entries about people; then all links to other people-articles contained in these entries. For each such link we calculate if the linked person was born after or before the subject of the article; if the life spans of the two overlap, then this connection is considered to be a likely personal acquaintance.

(4) Gravebook: Peers and influences

This is a feature of Gravebook that allows the user to view the "peers", i.e. personal acquaintances of the person of interest, the people who influenced this person, as well as those who have been influenced by this person. These three categories are determined by the Gravebook's social graph, using heuristics based on coccurence in Wikipedia and lifetime dates for each pair of people.

(5) Gravebook "Visualize peers" feature

For each person of interest, the Gravebook produces a visualization of his or her social networks, based on the "spring-box" algorithms that calculates the relative positions and distances between this person and the peers as represented in Wikipedia.

Figures 1 through 9 below illustrate different features of the interface. Figure 1 below shows the links to the articles on Johann Sebastian Bach found by the tool in different historical editions of Encyclopedia Britannica and in Wikipedia.

Appearance in Encyclopedia Britannica

Edition 9	Edition 11	Edition 15
BACH, JOHANN-SEBASTI	JOHANN CHRISTOPH FRIEDRICH BACH (1732...	Bach, Johann Sebastian

Figure 1: Encyclopedia pages for Johann Sebastian Bach.

Figure 2 shows the relative importance of articles on Bach across different editions, as measured by relative article volume, compared to the volume of the entire encyclopedia. Note that due to OCR noise, the current algorithm linked in an article on Bach's son in the 11th edition. Figure 3 shows the links to the articles on Voltaire in different historical editions of Encyclopedia Britannica and in Wikipedia. Figure 4 shows the relative importance of articles on Voltaire across different editions, as measured by relative article volume, compared to the volume of the entire encyclopedia. Figure 5 shows the relative importance of different members of the "English writers" category

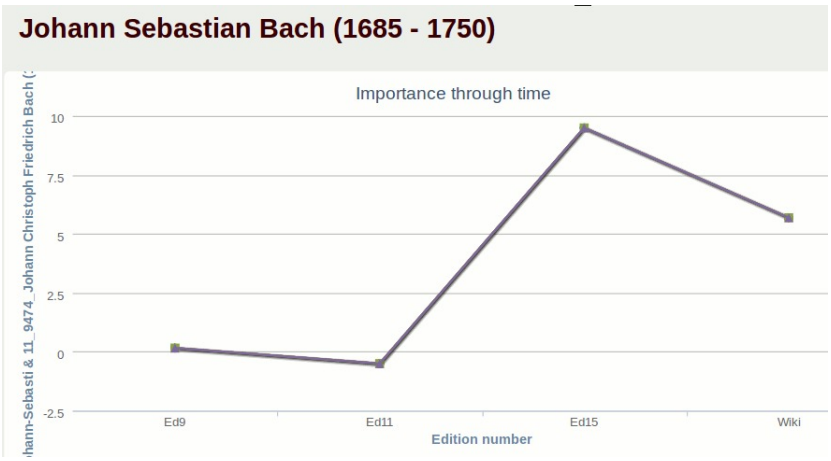
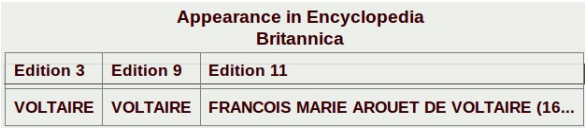


Figure 2: Importance of Johann Sebastian Bach as reflected in different editions.



Appearance in Encyclopedia Britannica		
Edition 3	Edition 9	Edition 11
VOLTAIRE	VOLTAIRE	FRANCOIS MARIE AROUET DE VOLTAIRE (16...

Figure 3: Encyclopedia pages for Voltaire

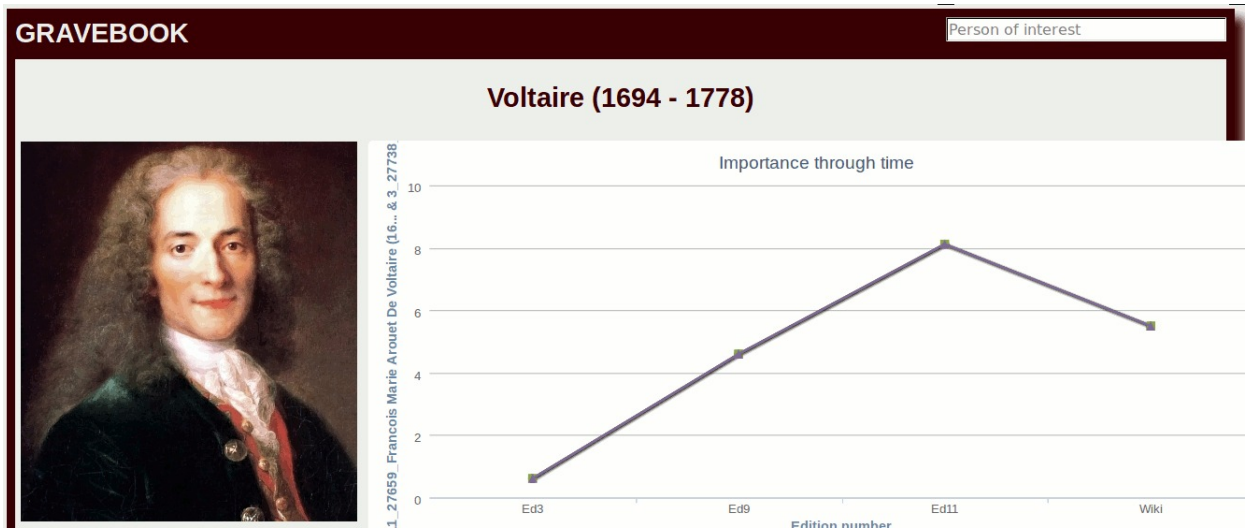


Figure 4: Importance of Voltaire as reflected in different editions.

through time. Importance for a category is computed in the same way as importance for its single member, treating the category as the union of the text in all member articles. Figure 7 shows peer communities of Bach's connections. The goal here is to visualize Bach's social circles. The interface allows the user to adjust the granularity of the social circle display, as seen in Figure 7. Figures 8 and 9 below show visualizations of social circles for Shakespeare and for Pushkin.

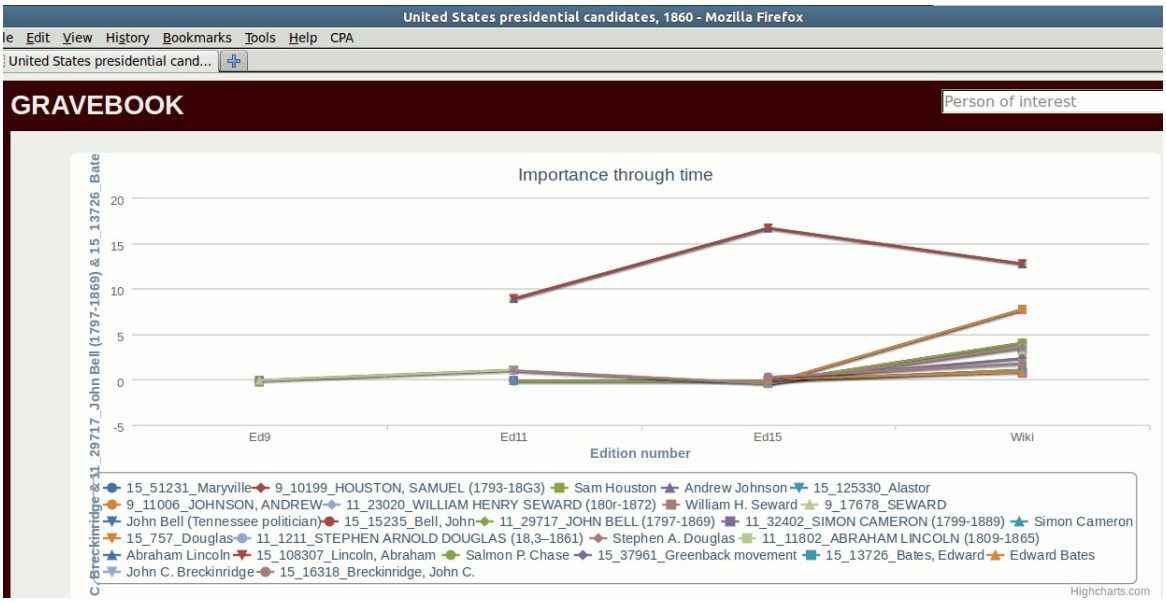


Figure 5: Importance of different members of the category of English writers through time.

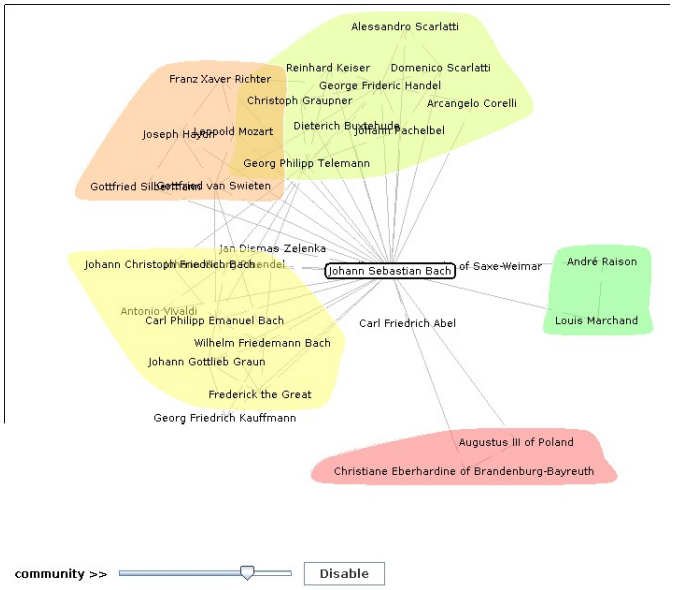


Figure 6: Visualizing Bach's social circles.

6.1.2 OCR noise correction spellchecker tool

In order to improve cross-edition article matching, we developed a noise correction spellchecker tool, which relies on the lexical information from the clean digitized edition of Britannica (Edition 15) in order to suggest corrections of commonly misspelled words. The tool runs over frequency dictionaries compiled from different editions. The list of corrections proposed by the tool is compiled by using an algorithm similar to such iterative methods of parameter estimation as the expectation maximization (EM) algorithm in statistics.

The tool implements the following algorithm:

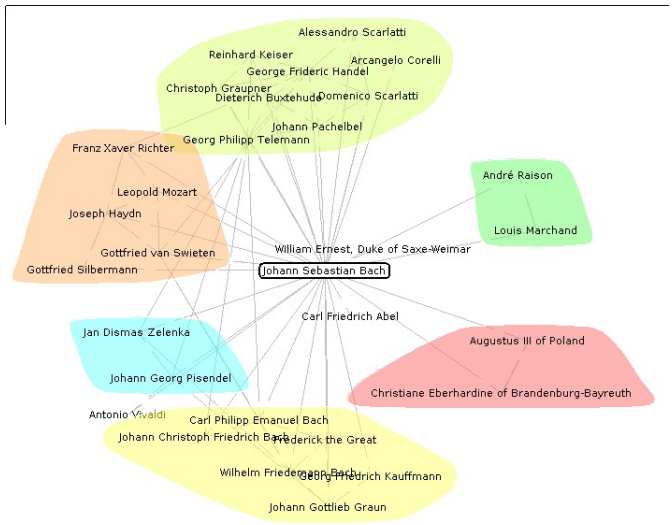


Figure 7: Visualizing Bach's social circles at a different granularity.

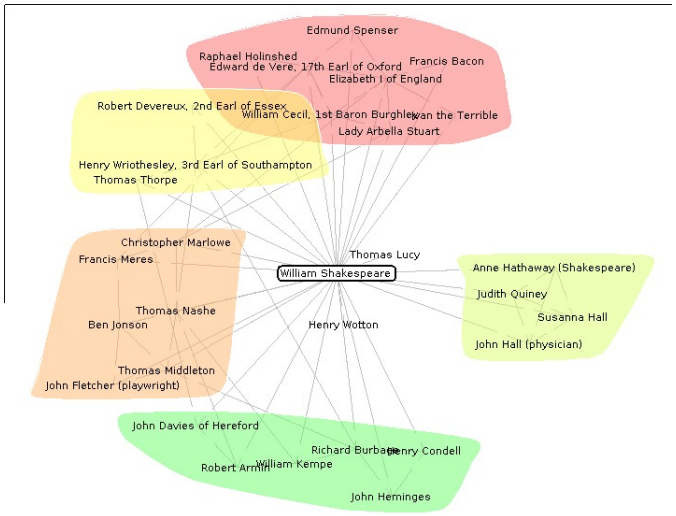


Figure 8: Visualizing Shakespeare's social circles.

1. Words that occur frequently in older digitized editions, but are absent from the latest edition of Britannica, are compiled into a dictionary of suspect words (likely errors), along with their frequencies across all the older editions.
2. Every word in the suspect dictionary is checked for whether it looks like a regular Arabic or a Roman numeral. If it does, then each letter is replaced with a number based on a pre-defined set of mappings: i $\rightarrow$ 1, o $\rightarrow$ 0, J $\rightarrow$ 5, etc. Results are saved directly to the replacement dictionary and no further changes are suggested for that word.
3. Any word longer than 15 characters is assumed to be a unique word, and no further changes are suggested for that word.
4. For every remaining word, every possible change is made up to an edit distance of 1. This could be set to 2, but any larger edit distances would produce inaccurate correction suggestions and would take too long, requiring a different algorithm. Changes are defined as inserting

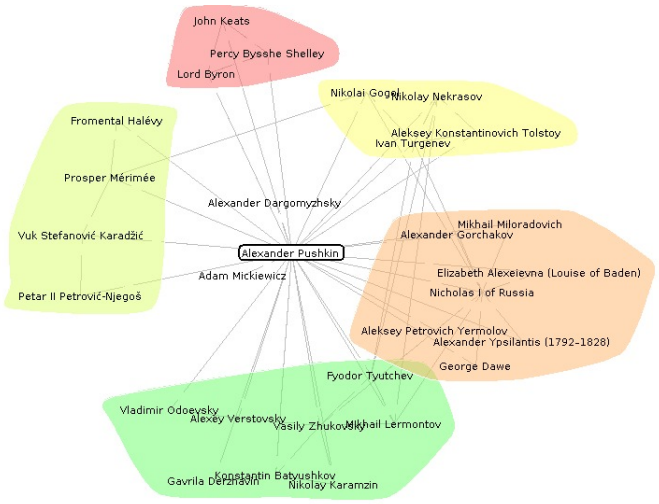


Figure 9: Visualizing Pushkin's social circles.

any character, deleting any character or substituting a character for another.

5. The resulting set of possible changes is then scored using the frequency of this change in the last iteration of the algorithm (or a default on the first iteration), divided by the frequency of the letter being changed to something else in the entire dataset (or a default for deletions). Insertions and deletions are underweighted, as compared to substitutions.
6. Every proposed change that results in a known word, is ranked based on the overall frequency of the resulting word times the score computed in step 5.
7. The top-rated change is selected. It is accepted as correct only if the score from step 6 is above a certain threshold AND the resulting word is more common than the suspected error by a certain factor. Otherwise, the original word is assumed to be correct, and no further changes are suggested for it.
8. The frequency of the original word is added to the score for the change made in step 5 to be used in the next iteration.
9. Once every word has finished processing, replace the list of change frequencies with the one computed in step 8, and start over at step 4.

Since steps 4 to 9 are repeated iteratively, the algorithm is much more accurate in the second iteration than the first, although it plateaus there. Note that this version of the algorithm includes no concept of which characters actually look like others – if this needs to be included, they could be entered in the first iteration for the frequency of a given change being made. The algorithm could easily produce a ranked list of changes rather than only one. All defaults and thresholds are accepted as command line arguments, and can thus be refined, or selected to be as accurate as possible by a testing harness.

In summary, the algorithm takes each word in the dictionary of suspect words, makes an initial guess to what the correct word is; counts the number of times it changed each character for each other, and then uses these counts to reprocess the words and make new guesses. If the word is not in the suspect dictionary, it is assumed to be unique and no corrections are proposed. Note that the algorithm is not context-sensitive. Also, a word may be assumed unique if it can only be modified by an improbable transformation, or a transformation would change a common error into an uncommon word.

This software is available as open source.

6.2 Publications

M. Gronas, A. Rumshisky, A. Gabrovski, S. Kovaka, H. Chen. Tracking the History of Knowledge Using Historical Editions of Encyclopedia Britannica in: LREC 2012, Workshop on Adaptation of Language Resources and Tools for Processing Cultural Heritage Objects.Proceedings. http://people.csail.mit.edu/arum/publications/GronasRumshiskyEtAlLrec2012.pdf

6.3 Course Materials

The instruments and methods developed within the grant project are currently being used as pedagogical tools and a case study in the course Introduction to Digital Humanities, at the department of Computational Linguistics, Higher School of Economics, Moscow, Russia

6.4 Collaborations

The project stirred interest in the newly formed digital humanities community in Russia, leading to a collaboration with the Center for the New Media and Society at the Moscow New Economic School, the first Russian institution for the study of digital culture and politics.[1]

Acknowledgments

We would like to acknowledge a wonderful team of students who worked on the project and without whose contributions this work would have been impossible. The project team included the following students:

- Aleksandar "Sasho" Gabrovski
- Hongyu Chen
- Matthew Digman
- Samuel Kovaka

[1] http://www.newmediacenter.ru/