

Call for Papers (DE)

Please find an English version below.
Veuillez trouver une version française ci-dessous.

#dhiha7: Wissensrohstoff Text.

Text Mining in den Geschichtswissenschaften

Mareike König (DHI Paris) – Marie Puren (INRIA Paris) – Matthias Lemke (DHI Paris)

Gegenstand der Tagung

„What do you do with a Million books?“ – Diese Frage Gregory Cranes reflektiert nicht nur ein für die Geistes- und Sozialwissenschaften zunehmend virulenter werdendes Erkenntnisproblem. Sie markiert gleichzeitig eine von der Altertumswissenschaft ausgehende methodologische Wende, die sich derzeit vollzieht. Sie besteht in dem Bemühen, zunehmend digital verfügbare, strukturierte und unstrukturierte Daten für die geistes- und sozialwissenschaftliche Forschung nutzbar zu machen. Nutzbar meint hier zweierlei: Es geht einerseits darum, Daten für die Forschung zur Verfügung zu stellen (Digitalisierung, Retrodigitalisierung) und andererseits darum, Verfahren anzubieten, um mit diesen potenziell sehr großen Datenmengen arbeiten zu können (Text Mining, Data Mining).

Text Mining wird bisher vor allem in der Philologie verwendet, zum Beispiel um Texte anhand von Stilmitteln wie Worthäufigkeiten oder gemeinsam auftretenden Worten einem bestimmten Autor oder Autorin zuzuschreiben oder um Gattungen, Werke und Epochen zu charakterisieren. Mit solchen semi-automatischen Analyseverfahren konfrontiert, beziehen sich viele HistorikerInnen auf die hermeneutische Tradition ihres Fachs und stehen Text Mining-Methoden zunächst eher skeptisch gegenüber. Das könnte sich in dem Maße ändern, indem immer mehr Texte online zur Verfügung stehen und indem bestehende Projekte mit Text Mining zur Beantwortung konkreter Forschungsfragen beitragen.

Um das Potenzial von Text Mining-Verfahren für die Geschichtswissenschaft nutzbar zu machen, bedarf es geeigneter Tools, die Frequenzanalysen, Kookkurrenzanalysen, Topic Modelling, Sentimentanalysen, Annotationsverfahren, Machine Learning und weitere Verfahren für historische Forschungsanliegen bereitstellen. Im Rahmen bereits abgeschlossener Projektförderungen existieren Infrastrukturen, Pilotprojekte etc., die diese und weitere Verfahren vorhalten. Ein großer Teil davon erlaubt zudem, die Verfahren einzusetzen, ohne dass die AnwenderInnen über Programmierkenntnisse verfügen müssen. Die Tagung ist von der Vorstellung getragen, dass das Einfordern von Programmierkenntnissen eine zusätzliche Hürde darstellt und den digitalen Methodeneinsatz verzögert, wenn nicht gar verhindert.

Die Ziele der Tagung bestehen daher zum einen darin, anhand ausgewählter Infrastrukturen Verfahren, Anwendungen und Analysestrategien für Text Mining vor dem Hintergrund konkreter Forschungsfragen vorzustellen und zu diskutieren. Zum anderen möchten wir anhand von einzelnen

Projekte zeigen, wie sich Text Mining in das Methodenarsenal von HistorikerInnen einbinden lässt und welche Antworten damit erzielt werden können. Dadurch soll für die geschichtswissenschaftliche Community die Bandbreite der Möglichkeiten zur analytischen Anwendung von Text Mining sichtbar werden. Da entsprechende Infrastrukturen, Tools und Projekte oftmals über disziplinäre und nationale Grenzen hinaus nicht bekannt sind, streben wir eine möglichst internationale Besetzung der Panels an.

Insbesondere sind wir an Beiträgen zu folgenden **Themenfeldern** interessiert:

- **Forschungsinfrastrukturen für Text Mining** (Funktionalität, Entwicklungsstand, Zugänglichkeit etc.)
- **Forschungsstrategien zur Anwendung von Text Mining** (Analyseschritte, Datenaufbereitung und -auswahl, Blended Reading etc.)
- **Vorstellung konkreter historischer Forschungsprojekte**, bei denen Text Mining in allen Facetten zur Anwendung kam (z.B. auch Sentimentanalyse)
- **Rechtliche Aspekte** (Urheberrecht, Zugang zu Korpora, Nachnutzung von Daten oder Korpora etc.)
- **Reflexion der epistemologische Änderungen der Geschichtswissenschaft** durch die Anwendung von Text Mining

Die Tagung ist als siebte Tagung in der vom DHIP seit 2011 durchgeführten, international viel beachteten Tagungsreihe „Digital Humanities am DHIP“ geplant. Die Tagungen werden mit wechselnden Partnern und Forschungsförderern durchgeführt und über ein deutsches und ein französische Blog medial begleitet. Die siebte Ausgabe richtet das DHI Paris in Kooperation mit dem INRIA, Institut national de recherche dédié au numérique, und der dort angesiedelten Forschergruppe ALMAnaCH aus. Das INRIA ist eine frankreichweit agierende Forschungseinrichtung mit mehreren Zentren im Bereich der Digital Humanities, die neben eigenen Forschungsvorhaben auch Infrastrukturprojekte durchführt.

Termin und Ort

Donnerstag, 29. Juni 2017, ab 14:00 Uhr, bis Freitag, 30. Juni 2017, 16:00 Uhr
Deutsches Historisches Institut Paris, 8, Rue du Parc Royal, F-75003 Paris, Frankreich
Tagungssprache: Englisch

Beiträge

Bitte schicken Sie einen Abstract (in englischer Sprache) im Umfang von maximal 300 Worten bis zum **28.2.2017** unter Angabe Ihrer Kontaktdaten an dhiha7@dhi-paris.fr.

Appel à communication (FR)

#dhiha7: Le texte comme matière première du savoir/ Le texte comme ressource (du/pour le savoir).

Le Text Mining pour les historiens

Mareike König (IHA Paris) – Marie Puren (INRIA Paris) – Matthias Lemke (IHA Paris)

Sujet/ objet du colloque

„What do you do with a Million books ?“ – Cette question, formulée par Gregory Crane, reflète non seulement un futur problème épistémologique - toujours plus prégnant - pour les sciences humaines et sociales, mais elle souligne également le tournant méthodologique, actuellement en cours et sur le point d'aboutir, des sciences de l'Antiquité. Ce dernier prend corps dans les constants efforts déployés pour rendre utilisables les données numériques disponibles – qu'elles soient structurées ou non – pour la recherche en sciences humaines et sociales. Le terme « utilisable » recouvre ici deux aspects : d'une part, il s'agit de mettre des données à disposition des chercheurs (grâce à la numérisation, et à la rétro-conversion) ; et d'autre part, il s'agit de proposer des méthodes afin de pouvoir travailler avec ces ensembles de données, potentiellement très importants (avec la fouille de textes et de données ou Text and Data Mining).

Jusqu'à présent, la fouille de textes a avant tout été utilisée par la philologie. Par exemple, afin d'attribuer des textes à un auteur précis en repérant des figures de style, comme les fréquences de mots ou l'apparition combinée de certain d'entre eux ; ou bien pour caractériser des genres, des œuvres ou des époques. Confrontés à de tels procédés d'analyse semi-automatiques, beaucoup d'historiens/ennes renvoient à la tradition herméneutique de leur discipline, et se montrent plutôt sceptiques à l'égard des méthodes offertes par le Text Mining. Toutefois, cela pourrait changer au fil du temps, tandis que de plus en plus de textes sont disponibles en ligne, et que des projets existants contribuent, avec la fouille de textes, à donner des réponses concrètes à des questions de recherche.

Pour que le potentiel des méthodes offertes par le Text Mining soit pleinement utilisé par les historiens/ennes, il faut mettre à disposition de la recherche en histoire des outils appropriés à l'analyse de la fréquence des mots, à l'analyse des cooccurrences, au Topic Modelling, à l'analyse des sentiments (ou Opinion Modelling), aux annotations, au Machine Learning,... et à bien d'autres procédés. Dans le cadre de financements indépendants, il existe déjà des infrastructures, des projets pilotes... etc, qui mettent en lumière ces méthodes, et d'autres encore. De surcroît, une grande partie d'entre eux permet d'avoir recours à ces méthodes, sans que l'utilisateur/trice n'ait besoin de connaître des langages de programmation. Le colloque est sous-tendu par l'idée qu'exiger des

connaissances de ce type constitue un obstacle supplémentaire, et retarde même l'application de ces méthodes, quand cela ne l'empêche pas.

Ce colloque a donc pour objectif d'explorer et de discuter, en s'appuyant sur les infrastructures sélectionnées, les méthodes, les applications et les stratégies d'analyse pour le Text Mining, à la lumière de questions de recherche concrètes. Par ailleurs, nous souhaitons montrer, à l'aide de quelques projets, comment la fouille de textes prend place au sein de l'arsenal de méthodes à disposition des historiens/ennes, et à quelles réponses on peut parvenir avec le Text Mining. De cette façon, la communauté historienne pourra avoir un aperçu clair de l'éventail des possibilités offertes par la fouille de textes. Comme les structures, les outils et les projets qui y sont dédiés restent souvent inconnus hors des frontières disciplinaires et nationales, nous sommes désireux de constituer autant que possible des panels de discussion avec une dimension internationale.

Nous souhaitons plus particulièrement recevoir des contributions s'intéressant aux thèmes suivants :

- **Infrastructures de recherche pour le Text Mining** (fonctionnalités, niveau de développement, accès, etc.)
- **Stratégies de recherche pour l'utilisation du Text Mining** (phase d'analyse, traitement et choix des données, Blended Reading, etc.)
- **Présentation de projets de recherche concrets en histoire**, qui présentent tous les facettes d'utilisation du Text Mining (par exemple, analyse de sentiments ou Opinion Mining)
- **Aspects juridiques** (droits d'auteur, accès aux corpus, reproduction des données ou des corpus...etc.)
- **Réflexion autour des changements épistémologiques en cours dans les sciences historiques** avec l'utilisation du Text Mining

Il s'agit du septième colloque organisé par l'IHA Paris depuis 2011 dans le cadre de la série de colloques, hautement reconnue au niveau international, intitulée « Les Digital Humanities à l'IHA ». Ces colloques sont organisés avec différents partenaires et institutions de financement de la recherche, et ils sont accompagnés par un blog franco-allemand. Cette septième édition est organisée par l'IHA Paris en collaboration avec Inria, et plus particulièrement avec l'équipe de recherche ALMAnaCH, spécialisée dans le traitement automatique des langues et les humanités numériques. Inria est un institut de recherche français de pointe, dédié aux sciences du numériques.

Informations pratiques

Jeudi 29 juin 2017 à partir de 14h et **vendredi 30 juin 2017** jusqu'à 16h

Institut Historique Allemand (Paris)

8, Rue du Parc Royal, F-75003 Paris, France

Langue du colloque: anglais

Envoi des propositions

Les propositions (en anglais) –300 mots maximum – sont à adresser à dhiha7@dhi-paris.fr avant le 28 février 2017.

Call for Papers (EN)

#dhiha7: Text as a resource.

Text Mining in Historical Science

Mareike König (GHI Paris) – Marie Puren (INRIA Paris) – Matthias Lemke (GHI Paris)

Key issues of the conference

“What do you do with a million books?” – Gregory Crane’s question does not only raise questions about the conditions in which the Humanities acquire knowledge. It can also be seen as indicating a methodological turning-point, unfolding today, coming from the discipline of Ancient History. It consists of all the efforts, trying to make the increasing amount of digital data, structured or unstructured, available for research in the Humanities. Making available in this sense has two sides: one is about making data accessible for scientists (e.g. digitization, retro-digitization), the other is about providing tools which allow to dig into the data, especially into huge amounts of data, for analytical purpose (Text Mining, Data Mining).

So far, Text Mining has mainly been used in philology, in order to – for example – attribute texts of a certain phrasing or word frequencies to a specific author, or to characterize typical language patterns of certain genres, eras or works. Being confronted with semi-automatic analysis, Historians, referring to the hermeneutical tradition of their subject, usually remain rather skeptical. This attitude might change, the more data for historical research is available and the more projects with a specific outcome have been accomplished, showing the value of Text Mining when it comes to answer certain research questions.

In order to unfold the potential of Text Mining within historical research, the availability of appropriate tools, providing and integrating frequency analysis, co-occurrence analysis, topic modelling, sentiment analysis, annotation procedures, machine learning and others, is crucial. Until today, research funding has already allowed to develop these tools and infrastructures, which exist and which are working. A great part of them also allows the use of Text Mining-tools without necessarily requiring a deeper or any understanding of programming or computer sciences. This is important, because our conference is mainly based on the assumption, that requiring competences in programming or computer sciences would create an additional barrier for Historians trying to use digital methods or could even prevent them from doing so.

Therefore, the conference has two major aims. One is about presenting existing infrastructures, tools, applications and analysis strategies for the use of Text Mining in specific research contexts. The other is about presenting research projects having applied Text Mining, which will allow us to show how to apply these methods and what outcome this could mean. These two aims taken together, the

conference will allow for researches in Historical Science to get an overview of existing techniques and strategies for mining huge amounts of text data. As infrastructures, tools and projects – given the predominant logic of national research funding – are most likely not known beyond their national scientific system, we also aim at bringing together a wide-range, international group of scientists in this conference, allowing exchange and cooperation.

Therefore, we are especially interested in contributions, which address the following **fields of research**:

- **Research infrastructures for Text Mining** (what functions are available, what is their current status, regarding development and accessibility, etc.)
- **Strategies for an appropriate application of Text Mining** (consecutive steps or parts of analysis, data processing, blended reading, etc.)
- **Presentation of research projects in any field of History**, having used Text Mining in various forms (e.g. sentiment-analysis or others)
- **Legal aspects** (copyright, access to data and corpora, re-use of data and corpora, etc.)
- **Reflections on epistemological changes in Historical Science** due to the application of Text Mining

This conference will be the seventh in the series of “Digital Humanities at the GHI”, which received great international attention and which has been presented by the Humanities Department of GHI Paris from 2011 on. The conferences have been and are organized with various partner and funding institutions. They also get a live coverage in a German and French speaking science blog. For this seventh conference, the GHI has the honor to cooperate with INRIA, the French Institut National de la Recherche dédié au Numérique, and its research team ALMAnaCH, which is specialized in the automated analysis of language.

Time and venue

Thursday, June 29th 2017 from 2p.m. and **Friday, June 30th 2017**, until 4p.m.

German Historical Institute (Paris)

8, Rue du Parc Royal, F-75003 Paris, France

Conference will be held in English.

Paper submission

Paper submissions (in English) –300 words max. – can be sent to à dhiha7@dhi-paris.fr until February 28th 2017.