

Open Research Online

The Open University's repository of research publications and other research outputs

XIP Dashboard: visual analytics from automated rhetorical parsing of scientific metadiscourse

Conference or Workshop Item

How to cite:

Simsek, Duygu; Buckingham Shum, Simon; Sandor, Agnes; De Liddo, Anna and Ferguson, Rebecca (2013). XIP Dashboard: visual analytics from automated rhetorical parsing of scientific metadiscourse. In: 1st International Workshop on Discourse-Centric Learning Analytics, 8 Apr 2013, Leuven, Belgium.

For guidance on citations see [FAQs](#).

© 2013 The Authors

Version: Accepted Manuscript

Link(s) to article on publisher's website:
<http://lakconference2013.wordpress.com/>

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

XIP Dashboard: Visual Analytics from Automated Rhetorical Parsing of Scientific Metadiscourse

Duygu Simsek¹, Simon Buckingham Shum¹, Ágnes Sándor²,
Anna De Liddo¹, Rebecca Ferguson³

¹ Knowledge Media Institute

³ Institute of Educational Technology

The Open University

Walton Hall, Milton Keynes, MK7 6AA, UK

{firstname.lastname}@open.ac.uk

² Parsing & Semantics Group

Xerox Research Centre Europe

6 chemin Maupertuis, F-38240 Meylan

France

agnes.sandor@xrce.xerox.com

ABSTRACT

A key competency that we seek to build in learners is a critical mind, i.e. ability to engage with the ideas in the literature, and to identify when significant claims are being made in articles. The ability to decode such moves in texts is essential, as is the ability to make such moves in one's own writing. Computational techniques for extracting them are becoming available, using Natural Language Processing (NLP) tuned to recognize the rhetorical signals that authors use when making a significant scholarly move. After reviewing related NLP work, we introduce the Xerox Incremental Parser (XIP), note previous work to render its output, and then motivate the design of the *XIP Dashboard*, a set of visual analytics modules built on XIP output, using the LAK/EDM open dataset as a test corpus. We report preliminary user reactions to a paper prototype of such a novel dashboard, describe the visualizations implemented to date, and present user scenarios for learners, educators and researchers. We conclude with a summary of ongoing design refinements, potential platform integrations, and questions that need to be investigated through end-user evaluations.

Categories and Subject Descriptors

K.3.1 [Computers and Education]: Computer Uses in Education

General Terms

Measurement, Design, Human Factors, Theory

Keywords

Learning Analytics, Natural Language Processing, Discourse, Metadiscourse, Argumentation, Rhetoric, Visualization

1. INTRODUCTION

A key competency that we seek to build in learners is a critical mind, i.e. ability to engage with the ideas in the literature, and to identify when significant claims are being made in articles. The ability to decode such moves in texts begins at school, and continues through secondary and tertiary levels, to doctoral research where learners must evidence their own capacity at a professional level. Understanding how to read (and ultimately write) according to the norms and criteria of a community is a core competency, with peer-reviewed publications serving as evidence of membership. Moreover, as literatures explode in size, and as fields become multidisciplinary, it is increasingly common for experienced researchers to find themselves navigating papers in unfamiliar communities, possibly written according to conventions different from their 'home' disciplines

(indeed Learning Analytics exemplifies such a field). For this reason, we are interested in analytics tools which may be able to help readers parse scholarly and scientific texts more rapidly, and which might provide feedback to students or to experienced researchers on the quality of their own writing. However, this paper restricts itself to making sense of the published literature.

In this paper we describe the discourse analysis module of the Xerox Incremental Parser (XIP) [26], which is of particular interest for learning analytics, since it identifies rhetorically salient sentences within scientific research papers. Built on this analysis we have designed visual analytics in the *XIP Dashboard*, which provides a range of summary overviews of the whole document corpus, rather than of one document at a time. Ultimately, we aim to assist the reader in assessing the current state of the art in terms of trends, patterns, gaps and connections in the literature.

The following sections will respectively introduce research into the automated rhetorical parsing (§2), the approach used by XIP (§3), the rationale for designing a XIP Dashboard (§4), and then the design process we have followed from a paper prototyping user study to implementation with user scenarios (§5). We then outline future work on refining the design, technical integration, and user evaluation studies (§6).

2. AUTOMATED ANALYSIS OF SCIENTIFIC DISCOURSE

In the past 50 years, scientific discourse has been conveyed primarily through journal and conference articles. Although other channels of scientific communication are fast evolving, the article continues to be the standard academically accepted channel for transmitting research results.¹ Since the appearance of electronic publication of scientific articles huge efforts have been put into machine processing to provide more effective navigation around the literature, and what is of particular interest to learning analytics, into more effective *communication* and *interpretation* of ideas.

The main line of Natural Language Processing (NLP) research in scientific publication aims at extracting factual information from the texts of the articles and transforming them into structured data that can populate ontologies or databases (see

¹ Cf. "The number of scientific articles indexed by Thomson Reuters increased from fewer than 600,000 in 1990 to more than 1 million in 2009." (Times Higher Educ: <http://bit.ly/TES-ExpandingLiterature>)

e.g. [1-2]). Factual information extraction consists in extracting names and terms relevant for the domain and ontological relationships that hold among them. In the framework of factual information extraction each piece of extracted information is an entry in a data structure.² Scientific research, however, as we are arguing in this paper, does not consist in providing a list of facts, but it essentially consists in reasoned argumentation around facts. In the articles that account for their research the researchers make hypotheses, they support, refute, reconsider, confirm, build on previous ideas in order to support their own ideas and findings. Consequently the automatic processing of research articles should be able to capture and represent the evolution of ideas and findings, as they are described in the papers (for detailed argumentation see [3]).³

Research articles conform to rhetorical writing conventions that support the argumentative texture of the article and at the same time guide the reader in following it. The importance of these conventions, and thus the importance of rhetoric for composing research articles, is testified by the huge body and importance of literature describing the principles and techniques for writing research articles (e.g. [4-5] for a comprehensive picture).

A recently evolving direction in NLP considers these rhetorical practices as the basis for extracting information embedded into the discursive, argumentative, rhetorical nature of the research articles. The knowledge items thus extracted are labeled according to their rhetorical status in the article: aim, result, conclusion, new knowledge, old knowledge, open question, etc. This labeling allows further processing in a nuanced way. Among other traditional applications like summary writing or information retrieval, automated rhetorical annotation can also assist curators to populate semantically structured knowledge bases by pointing at hypotheses and claims, or can provide input to argumentative social network systems [31].

In order to illustrate the role of rhetorical development and argumentation in the constitution of knowledge conveyed in the research article, we present the first sentences of an abstract in biology. Factual information is in plain text, and rhetorically oriented expressions – often referred to as metadiscourse – are italicized.

Most evolutionists agree to consider that our present RNA/DNA/protein world has originated from a simpler world in which RNA played both the role of catalyst and genetic material. Recent findings from structural studies and comparative genomics now allow us to get a clearer picture of this transition. These data suggest that evolution occurred in several steps, first from an RNA to an RNA/protein world (defining two ages of the RNA world) and finally to the present world based on DNA.

² For example, and of direct relevance to scientific communication, there has been significant recent interest in shared formalisms for mapping elements of formal ontological structures to scientific or other documents on the web and in other formats such as PDFs, using models such as the Annotation Ontology [6].

³ Readers may also be interested in the use of sentiment mining techniques to detect support and opposition in product reviews or transcripts (e.g. political debates: [7]). These have not, to our knowledge, been applied to specific characteristics of scientific discourse, but represent an emerging field of potential relevance. Another line of work concerns modeling arguments in procedural texts, which provide the backing for following instructions (e.g. [8]), which again, might be relevant to scientific argumentation with respect to the rationale for following a particular methodology.

Discourse-oriented automated processing consists in the identification of the italicised elements – often referred to as metadiscourse – and their interpretation in terms of rhetorical functions. The following are some examples of discourse moves relevant in research articles conveyed by metadiscourse.

“Summarizing” is a function by which the author can refer to the issues dealt with in (parts of) the text. In the following examples the parts of the sentence carrying out the rhetorical function of “summarizing” are in italics:

The purpose of this article is to develop the idea that the procedures in any given classroom or laboratory exercise should be definitely determined by the specific aim, which the instructor has in mind to accomplish.

The perspective I shall use in this essay relies heavily on the view of professionalization presented in Andrew Abbott's brilliant study, *The System of Professions* (Abbott, 1988).

This paper explores social practices of propagating ‘memes’ (pronounced, ‘meems’) as a dimension of cultural production and transmission within Internet environments.

Authors carry out the rhetorical function of “contrasting ideas” when they contest, question or point out as significant or new some issues, facts or theories, when they indicate a gap in knowledge, or point out any flaw or contrast related to the subject, etc. In the following examples, the parts of the sentences carrying out the rhetorical function of “contrasting ideas” are in italics:

With an absence of detailed work on masculinities and sport in South African primary schools (for an exception, see Bhana, 2002) *this paper goes some way towards addressing the issues around young boys’ developing relationship with sport.*

My interest of inquiry emerged in 1997 from a new idea in school pedagogy and sport pedagogy.

Sentences conveying contrasting ideas may be further categorized into subclasses like *novelty, surprise, importance, emerging issue and open question.*

The automated extraction of such metadiscourse requires the identification of the italicized elements and their interpretation in terms of rhetorical functions. This is a difficult task for two main reasons:

- There exists a great variety of discourse and rhetorical models with various analysis units and goals.
- It is notoriously difficult to map linguistic expressions into argumentative and rhetorical moves, since human languages do not provide special resources dedicated to rhetorical functions.

Owing to these reasons, the few existing computational linguistics applications to the rhetorical analysis of scientific articles⁴ do not approach research articles through particular discourse linguistics models, but rather, propose robust discourse annotation methods inspired by a variety of models, while they rely on corpus analysis and are motivated by application needs.

⁴ There exists a great body of work in computational discourse analysis (e.g. [9-11]), however their categories and methods have not been applied to robust processing, which is required in information extraction tasks.

Table 1: Categories from Teufel’s Argumentative Zoning

BACKGROUND	Generally accepted background knowledge
OTHER	Specific other work
OWN	Own work: method, results, future work
AIM	Specific research goal
TEXTUAL	Textual section structure
CONTRAST	Contrast, comparison, weakness of other solution
BASIS	Other work provides basis for own work

2.1 Argumentative Zoning

In this section we present an influential approach to the automated analysis of scientific articles, Argumentative Zoning, before introducing our approach using the Xerox Incremental Parser (XIP). For other work in this direction see [12-15].

Argumentative Zoning was developed in the doctoral research of Simone Teufel [16]. This was the first attempt to automatically annotate rhetorical moves in research articles. Teufel establishes the Rhetorical Document Profile (RDP), which is “designed to encode typical information needs of new readers in a systematic and structured way”. As we see, the emphasis here is on “information needs” and not on “information”. The task is to automatically identify the parts of the articles that serve these information needs, which Teufel’s calls the *argumentative zones*.

Argumentative zones, which cover the entire article, are defined in terms of a “model of prototypical scientific argumentation” containing the argumentative moves in Table 1. This list is inspired and motivated by a variety of approaches to the analysis of scientific discourse. From a discourse analysis point of view it draws on Swales’ model of argumentative moves [17], and Hyland’s system of the description of metadiscourse [18]. From a practical point of view it aims at fulfilling requirements for detecting the attribution of intellectual ownership, citations and author stance. It also applies work on problem solving processes (e.g. [19-20]), and on the strategies of scientific argumentation [17], [21].

Argumentative zoning is described as a difficult task both from the point of view of the establishment of a gold standard for annotation and from the point of view of automated execution. Teufel concludes that new evaluation methods are required, since the interpretation of the results in terms of recall and precision is not straightforward either.

Originally, argumentative zoning was proposed for automatic summarization and information retrieval tasks. Later it was also used for educational purposes [22] and citation indexing [23]. Since the theory and technique of argumentative zoning are shown to be robust and operational, subsequent work consisted in annotation experiments in different disciplines, including chemistry [24] and biology [25].

3. XEROX INCREMENTAL PARSER

Sharing the basic assumption of Argumentative Zoning – that rhetorical moves can be detected from the author’s language use

Table 2: Rhetorical functions identified by XIP

SUMMARIZING	summarizing aims, claims, results, conclusions
BACKGROUND KNOWLEDGE	descriptions of previous ideas
CONTRASTING IDEAS	descriptions of ideas as contrasting
NOVELTY	descriptions of new ideas
SIGNIFICANCE	descriptions of ideas as being significant
SURPRISE	descriptions of ideas as being surprising
OPEN QUESTION	descriptions of open questions
GENERALIZING	descriptions of research trends

– a different approach is taken by the Xerox Incremental Parser (XIP) [26] for the rhetorical analysis of scientific articles. Instead of covering the whole article, this approach aims at highlighting the main research issues that the articles handle.

XIP annotates the rhetorical functions in Table 2 as bearing the main research issues. This choice of the rhetorical moves is motivated by various considerations. SUMMARIZING and BACKGROUND KNOWLEDGE relate to conveying main ideas in a straightforward way in the rhetorical construction of research articles. The other categories have their roots in Thomas Kuhn’s view of science as primarily a problem-solving activity [27]. Thus the *raison d’être* of any research paper is the problem, and the main ideas are to be found in sentences where the research issues are described. These sentences fulfill rhetorical functions of contesting, questioning or emphasizing research-related ideas, facts or theories as being significant or new research-related ideas, facts, or theories, of indicating a gap in knowledge, or of pointing out any flaw or contrast related to the research topic. This approach does not claim to provide a complete characterization of the research problem, neither does it represent the rhetorical construction of the article, but its main goal is to provide assistance in rapidly gaining understanding about the approach of the article to the research in question.

The rhetorical functions detected by XIP partly overlap with the argumentative zones, and partly are different from them. The main difference is that the contrasts among ideas are not approached from the point of view of intellectual ownership, but rather from the point of view of the various ways in which contrasting ideas are introduced.

There have been a number of proof-of-concept applications that justify the choice of these categories as bearing salient ideas:

1. Detecting abstracts in the PubMed database that describe substantially new findings [28]
2. Improving information retrieval in a search engine dedicated to educational science [29]
3. Reading assistance for peer-reviewers [30]
4. Analysing research project reports [31]

4. XIP DASHBOARD RATIONALE

As is typical of language technologies, XIP generates a semantically tagged file suitable for subsequent machine analysis (Figure 1).

```
<body>SUMMARY(Our work is a step towards building taxonomies based on both expert and data - driven approaches which we believe could lead to a needed trade-off between power and accuracy .  
  
<body>EMPHASIS(The classic categorization of Austin [14] postulates five major speech act classes based on five categories of performative verbs :  
  
<body>CONTRAST(A drawback of the proposed model for representing dialogue utterances , i.e. the N leading tokens , is that the distance between two dialogue utterances is based on string operations rather than lexico - semantic distances which would be more meaningful for natural language dialogues .
```

Figure 1: Raw XIP output showing sentence classification

In addition, other parsing modules in XIP extract concepts (nouns and noun phrases) from these sentences, which makes possible the prioritization of those concepts, on the grounds that they appear in contexts judged to be rhetorically significant (Figure 2).

```
CONCEPT(taxonomy)  
CONCEPT(effective tutoring dialogue)  
CONCEPT(need for reliable annotation)  
CONCEPT(emphasis on power)  
CONCEPT(interactions between the experts' conceptual view)  
CONCEPT(human annotators)
```

Figure 2: Raw XIP output showing extracted concepts

While plain textual output is well suited for researchers to analyse manually, or with other tools, this is not a form which could be usefully or attractively presented back to either learners, educators seeking to assess their progress, or to other kinds of information analyst for whom this work is relevant. In prior work, we have rendered XIP's output in two ways, as part of our explorations into combining human and machine annotation for "contested collective intelligence" [31]:⁵

- *Document-centric*: as annotations superimposed on the HTML version of the document, using a Firefox plug-in for Cohere, a social web annotation and knowledge mapping application;
- *Network-centric*: as nodes added to a self-organising network visualization of semantically classified nodes and edges.

These representations provide quite detailed 'zoomed in' views of XIP's output, but do not provide effective summary overviews which would allow a reader to grasp the overall quantity and quality of XIP's analysis, in order to choose where to inspect more closely. These kinds of visual analytics

dashboards are becoming increasingly available to educators (and to a lesser degree to students) in online learning platforms. The XIP Dashboard is being conceived as a suite of visual analytics on XIP output, to help draw attention to candidate patterns of potential significance to students, educators and experienced researchers alike:

- the occurrence of domain concepts in different metadiscourse contexts (e.g. effective tutoring dialogue in sentences classified contrast).
- trends of the above over time, e.g. to show the development of an idea
- trends within and differences between research communities, as reflected in their publications

5. XIP DASHBOARD DESIGN PROCESS

5.1 Paper prototype evaluation

Our assumption is that at least in the first instance, this novel form of analysis will not be a 'walk up and use' interface, but quite a specialized tool that will require some introduction, akin to the kinds of Advanced Search interfaces that most people do not use by default. Future versions will, we hope, mature into increasingly usable forms, but at present, the objective is not to let interaction design flaws obscure insights into initial user reactions to the concept of such a novel tool.

Prior to any software implementation, preliminary design work was done using a multi-screen paper prototype. The methodology was paper prototyping [32], which has been used successfully for low-fi storyboarding of user interfaces to elicit rapid feedback on novel interactive applications, and to get user data before coding.

Six user sessions were conducted with first year PhD students who were just a few months into their literature analyses at the Open University. Each session took around 45 minutes and consisted of three phases: testing the user interface and getting opinions about what might have been changed in the design, discussing how such a tool might address the problems users were already facing in conducting their literature reviews, and discussing if this tool could be used to analyze one's own writing.



Figure 3: XIP Dashboard paper prototype for user pilot

Sessions started with a brief introduction to the concept of a machine identifying specific kinds of sentences in papers, and a guided tour to the interface about how dashboard works (analogous to an instructional movie). Participants were then

⁵ See also the seminar and demonstration: <http://bit.ly/CohereXIP>

given two tasks, using a think-aloud protocol, and using their fingers as a mouse to ‘click’ on the screens, which the researcher would then change. All participants managed to navigate the interface and complete the tasks, providing preliminary feedback that the basic information architecture of the prototype was reasonable. There was consensus around switching from stacked, multi-variable bars in the charts, which they found hard to understand, to a more conventional set of single-variable columns in a bar chart.

When participants discussed the challenges they faced in their literature reviews, they all expressed enthusiasm about the idea of such a dashboard. Their intuitions were that this could save time by identifying more effectively the key papers and key sentences within papers. There were similarly positive reactions to the concept that the future aim of the dashboard might provide instant feedback on their own writing, such as blog posts and draft papers. They could envisage seeing more clearly how their writing was changing over time in terms of the rhetorical classifications, especially at this early stage in their research careers, as first year PhD students.

To summarise, one cannot read too much into preliminary user reactions to paper storyboarding of this sort, in which it is clearly impossible to go into real depth on tasks. However, we took encouragement from the positive reactions to such a novel way of reading and writing.

5.2 Document corpus and implementation

We selected the LAK Dataset published by the Society for Learning Analytics Research, which provides machine-readable plaintext versions of the proceedings of the *Learning Analytics and Knowledge (LAK)* conference and a journal special issue, and the *Educational Data Mining (EDM)* conferences and journal.⁶ All of the papers (66 LAK and 239 EDM papers, 305 in total) were analysed using XIP, extracting 7847 sentences and 40163 concepts.

The output files, one per paper, were imported into a MySQL database, and the user interface implemented using PHP and JavaScript, making use of Google Chart Tools for the interactive visualizations.⁷

5.3 Dashboard user interface

The dashboard consists of three tabs, each showing different analytical results in different types of charts. The first section consists of two line charts each representing the distribution of rhetorical sentences by year in LAK and EDM. Line charts depict sentences by rhetorical type over time (Figure 4). This provides a birds-eye-view of the distribution of rhetorical moves per year, with both literatures remaining stable for most types, but a clear separation in frequencies between relatively high and rising levels of *Summary* and *Contrast* moves compared to the others.

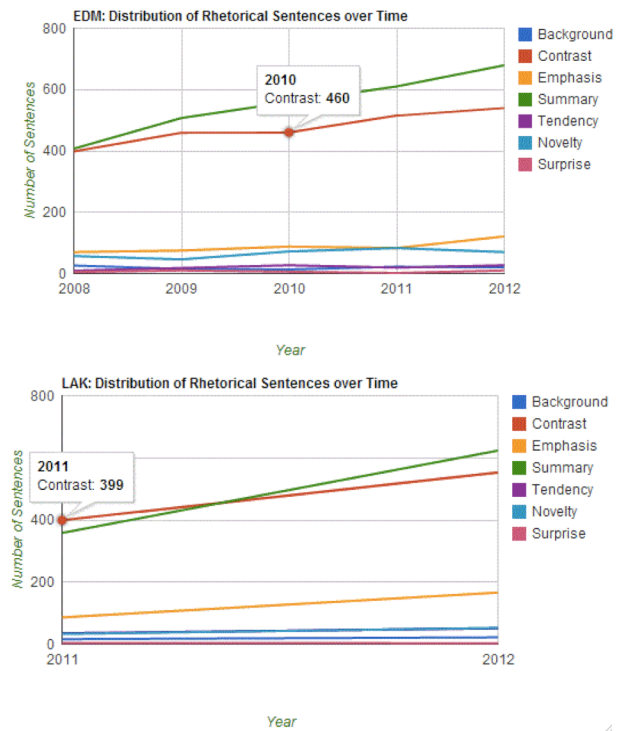


Figure 4: Rhetorically classified sentences graphed by frequency and year, for EDM and LAK

The next visualization permits users to specify a combination of the extracted concepts in which they are interested, to see the occurrence of these in papers within all or specified communities (Figure 5).

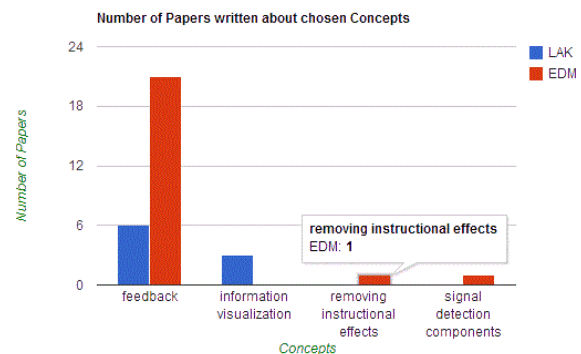


Figure 5: Number of papers with rhetorically extracted sentences containing user-selected concepts.

Thirdly, a bubble chart displays the occurrence of papers within any or all communities, filtered by user-selected concepts and year of publication (Figure 6). As shown by the colour spectrum at the top, saturation represents the ‘density’ of the concept in the paper, as defined by number of XIP classified sentences in which it occurs (where darker = denser).

⁶ LAK Dataset: <http://www.solaresearch.org/resources/lak-dataset>

⁷ Google Chart Tools: <https://developers.google.com/chart>

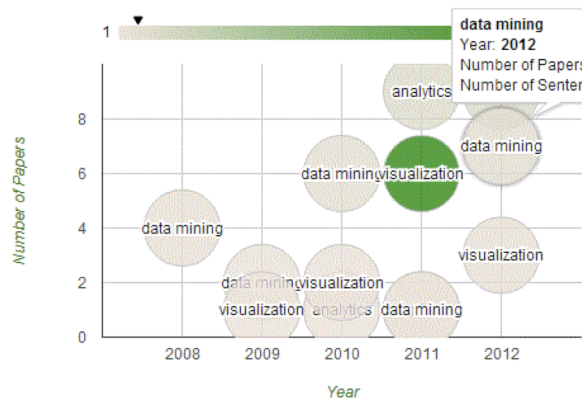


Figure 6: Concept 'density' within XIP sentences, by year and number of papers

There is no rhetorical analysis shown at this level of detail, but mousing on a concept bubble displays a pie chart showing the relative distribution of rhetorical types (Figure 7).

To summarise, we have described the first design iteration of a set of visual analytics which make visible, for the first time, the output of XIP text classification. This is a significant advance on working with hundreds of text files (Figures 1 and 2), which provided no cognitive support to the analyst in seeing any macro-level patterns in a corpus, or navigating between papers sharing common rhetorical moves and concepts.

The next step is to begin evaluating the potential of such views for end-user communities. The first step we can take is to describe candidate use scenarios.

5.4 Student/Educator/Researcher Scenarios

In order to better demonstrate our vision for the XIP Dashboard, consider the following user scenarios for a student, educator and experienced researcher.

5.4.1 Student Scenario: preparing for an essay

Jane is a 1st Year Digital Marketing master's student who is enrolled in a Web Analytics module. Part of her assessment requires her to write an essay about *How to Get into Big Data Analytics in Online Marketing*.

As Jane is quite new to this research area, her primary aim is to do a literature review. She's curious to know which community writes more on these topics, so she logs into her university library, which provides subscription access to e-journal and conference papers. She finds the LAK and EDM literature, and then switches to the rhetorical dashboard view, and selects the second tab (Figure 5). This allows her to specify a combination of key concepts related to her research, to see the occurrence of these in papers. She chooses *big data* and *data analytics* from the list of concepts and selects all communities to learn which community tends to publish more on these topics.

Once Jane gets the general overview about where to look, her next step is getting know the distribution of these numbers by year. Therefore she moves to the third tab (Figure 6), which displays the occurrence of papers within all communities, filtered by her selection of concepts and year of publication. This shows her when *big data* and *data analytics* began to referred to, in which years this peaked, and the overall trend.

Information Visualisation (2011): 3 papers, 6 sentences

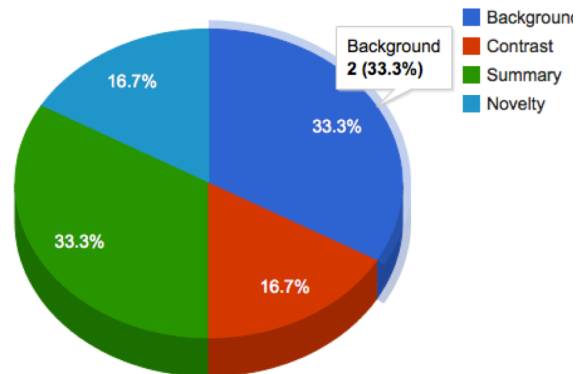


Figure 7: Distribution of rhetorical types in the sentences behind a selected concept bubble (Figure 6)

She picks a peak year, and wants to find contextualizing statements about the background of the topics, for her literature review, so she switches to examining the distribution of rhetorical types (Figure 7). She chooses sentences classified as *Background* topics, and from there, finds the paper listings.

5.4.2 Educator Scenario: assessing essays

Professor Jones is reviewing progress in her advanced level class taking *Educational Futures*. By this stage, the students should be capable of writing coherently structured essays with a clear thesis, backed by good argumentation, appropriately contextualized to the literature. She brings up the XIP Dashboard and points it to the folder with 45 essays, each 20 pages. A few seconds later, the visualizations have loaded, and she begins to explore. She can view this year's essays graphed against the preceding years 2011-2013, giving her a sense of whether there has been an overall change in the use of appropriate concepts, or writing style, but this doesn't seem to be the case. She's a bit annoyed about this, since she's been trying to improve her teaching of scholarly writing. Maybe it's just the students.

Drilling down to individual 2014 essays (Figure 7), she can see that for lower achieving students, the balance of rhetorical moves is quite skewed, with *Background* and *Summary* contributions dominating. While using these in the expected introductions and conclusions of their essays, her higher achieving students seem to make stronger, more assertive moves in which they *Contrast* claims, express *Surprise* about certain trends (associated with the concepts *MOOC* and *accreditation*).

Prof. Jones finds this so compelling that she applies for university ethics board approval to run a pilot with the 2015 cohort, to see if dashboard feedback on essay drafts proves useful.

5.4.3 Experienced Researcher Scenario

Dr. Mark Holmes is a senior lecturer whose research interests are focused on *learning analytics* and *data visualization*. He has been commissioned by the Department of Education to compile a report on *The State of Learning Analytics and Educational Data Mining in 2014*.

He wants to check his intuition that the EDM community writes far more about the quality of data mining algorithms and less about the end-user experience, while LAK puts more emphasis

on the sensemaking that goes on around the outputs of the algorithms. He uses the second and third tabs of the XIP dashboard (Figures 5-6) to explore the relative presence and trends of the concepts *algorithm*, *dashboard*, *visualization*, and *end-user evaluation* over the period 2011-2014. This confirms his hunch, and helps him identify clusters of papers thematically.

More interestingly, he can see a relative higher frequency of *Novelty* and *Contrast* sentences in EDM papers mentioning *algorithms* (and a set of specific algorithm names he has selected), suggesting that EDM research and argumentation goes into more depth than LAK on those issues.

6. FUTURE WORK

We have motivated the first implementation of the *XIP Dashboard*, but clearly much work remains, both short-term design refinements of current version, and to realize our future aim to test the dashboard as formative feedback for one's own writing.

6.1 Design refinements

Following a design review with the lead XIP analyst, and range of refinements are under way.

The remaining step from Figure 7's pie chart is to take the user from a slice of the pie chart to the corresponding list of XIP extracted sentences. Selecting one of these will then take the user to the full text showing the highlighted sentence in context.

The Figure 6 bubble chart will add one more dimension, using the size of the bubble to show how many papers are involved. The y-axis will now indicate the percentage of total number of papers rather than raw numbers, to correct for different sizes of corpus (Figure 8).

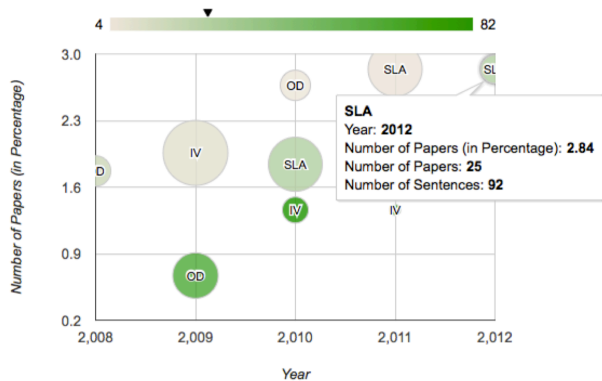


Figure 8: Mockup for version 2 of the bubble chart, using size of bubble to convey number of relevant papers

Beyond these refinements, a new visualization in design is a heat map that combines all concepts and all rhetorical markers in the same view. As the mockup in Figure 9 indicates, each cell displays from yellow to red the density of the concept and rhetorical classifications of the sentences in which it occurs. Mousing over the cell will display the matching sentences, and from there, the full text.

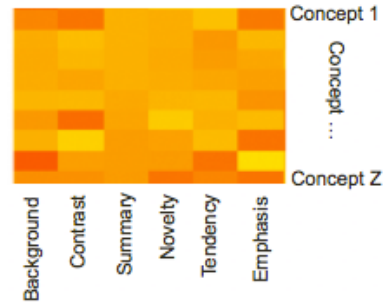


Figure 9: Mockup of interactive heat map showing the intersection of all rhetorical classifications and concepts

6.2 XIP integration with learning platforms

A longer-term goal is to build learners' capacity to read and write more like scientists — but scientific discourse is no longer restricted only to formal publications, central though these are. Beyond the institutionally sanctioned virtual learning environments, is the infinite expanse of web resources, and of particular interest from a DCLA perspective, the expanding sphere of social learning applications in the cloud, including blog posts, discussion forums, wikis, and social networking sites. Prior work has already shown that XIP can return useful results from analysing online discussion forums, and enhanced search results that go beyond conventional digital libraries [29]. We will be experimenting with invoking XIP as a web service from plug-ins that we develop on the experimental DCLA platforms that we are developing within the Knowledge Media Institute, such as Cohere [33], EnquiryBlogger [34] and the Evidence Hub [35].

6.3 End-user evaluation

As we put in place the different elements of the technical platform, we are approaching the point where we can begin evaluations. While user-centred design is important, learner-centred design requires that we focus on the pedagogical benefits. Beyond that, a Learning Analytics perspective demands that we also take into account the educators who may be using such analytics to prioritise their attention and assess student progress, and researchers who are studying the processes (cf. the scenarios).

We propose that this line of work opens up an interesting and important set of questions for future research into the impact of discourse analytics in learning (not just for XIP, but any automated annotation tool proposed for learners), such as:

- What is the signal to noise ratio from XIP, and does the XIP Dashboard have a role to play in personalized filtering? Does XIP classify a sufficiently useful proportion of sentences that users (who as noted above, vary in role) do not feel they are wasting their time? Can XIP Dashboard provide way for users to tune it to their personal interests?
- Do the shortcuts to reading, which XIP-annotated papers and corpora provide, in fact supplant the learner's own cognitive work, meaning that they do not engage as deeply? Or does automated annotation enable them to spend their time more effectively, and go deeper on relevant papers than when their attention is more thinly spread trying to identify them in the first place?

- Do educators and researchers find that XIP Dashboard provides qualitatively new insights into a document or cluster, or just helps to speed up existing assessment practices?
- How do writers respond to XIP formative feedback? Can we validate this feedback with other more conventional measures of writing quality? Do writers of different abilities – from school pupils, to undergraduates, to citizen scientists, to doctoral candidates, to professors – differ in the benefits they gain? Anecdotal evidence suggests that even experienced researchers can benefit from XIP, but this has yet to be systematically evidenced.

7. ACKNOWLEDGMENTS

We gratefully acknowledge funding from the Wolfson Foundation, whose doctoral studentship supports the first author as part of The Open University's OpenScience Lab. We also thank Xerox Research Centre Europe for enabling this collaboration with XIP.

8. REFERENCES

- [1] Corney, D., Buxton, B., Langdon, W., and Jones, D. (2004). BioRAT: extracting biological information from full-length papers. *Bioinformatics*, 20(17): 3206, 2004.
- [2] Garten, Y. and Altman, R. (2009). Pharmspresso: a text mining tool for extraction of pharmacogenomic concepts and relationships from full text. *BMC bioinformatics*, 10(Suppl 2): S6, 2009.
- [3] de Waard, A. and Pandermaat, H. (2009), Categorizing Epistemic Segment Types in Biology Research Articles. *Workshop on Linguistic and Psycholinguistic Approaches to Text Structuring (LPTS 2009)*, September 21-23 2009. – to be published as a chapter in *Linguistic and Psycholinguistic Approaches to Text Structuring*, Laure Sarda, Shirley Carter Thomas & Benjamin Fagard (eds), John Benjamins, (planned for 2010).
- [4] Swales, J. (1990). *Genre Analysis: English in Academic and Research Settings*. Cambridge, UK: Cambridge University Press.
- [5] Hyland, K. (2005). *Metadiscourse*. Continuum, 2005.
- [6] Ciccarese, P., Ocana, M., Chalfont, C., Dow, E., West, A., and Clark, T. (2010). The SWAN Annotation Framework. *Manuscript in preparation*.
- [7] Thomas, M., Pang, B., and Lee, L. (2006). Get out the vote: Determining support or opposition from Congressional floor-debate transcripts. *Proceedings of Empirical Methods in Natural Language Processing (EMNLP)*, pp. 327–335, 2006
- [8] Nordström, B. and Ranta, A. (2008). Analyzing Argumentative Structures in Procedural Texts. *Advances in Natural Language Processing, 6th International Conference*, Gothenburg, Sweden, August 25-27, 2008. Lecture Notes in Computer Science, Volume 5221, 2008. Berlin: Springer
- [9] Marcu, D. (2000). *The Theory and Practice of Discourse Parsing and Summarization*, MIT Press, Cambridge, MA, 2000.
- [10] Mann, W. C. and Thompson, S. (1987). Rhetorical Structure Theory: Toward a functional theory of text organization. *Text*, 8 (3): 243-281.
- [11] Polanyi, L., Thione, G.L., van den Berg, M., and Culy, Ch. (2004). LiveTree: An Integrated Workbench for Discourse Processing (2004) ACL2004 - *Workshop on Discourse Annotation*
- [12] Pendar, N. and Cotos, E. (2008). Automatic identification of discourse moves in scientific article introductions, *Proceedings of the Third Workshop on Innovative Use of NLP for Building Educational Applications*, p.62-70, June 19-19, 2008, Columbus, Ohio
- [13] Ruch, P., Boyer, C., Chichester, Ch., Tbahriti, I., Geissbühler, A., Fabry, P., Gobeill, J., Pillet, V., Rebholz-Schuhmann, D., Lovis, C., and Veuthey, A-L. (2007a). Using Argumentation to Extract Key Sentences from Biomedical Abstracts, *Int J Med Inform*, 76(2-3): 195-200. 2007.
- [14] Ruch, P., Geissbühler, A., Gobeill, J., Lisacek, F., Tbahriti, I., Veuthey, A-L., and Aronson, A.R. (2007b). Using Discourse Analysis to Improve Text Categorization in MEDLINE. *Medinfo* 2007.
- [15] Burstein, J.C., Marcu, D., and Knight, K. (2003). Finding the WRITE stuff: automatic identification of discourse structure in student essays. *IEEE Intelligent Systems*, 18(1): 32–39.
- [16] Teufel, S. (1999). Argumentative Zoning: Information Extraction from Scientific Articles. *Doctoral Dissertation*, University of Edinburgh
- [17] Swales, J. (1990). *Genre Analysis: English in Academic and Research Settings*. Cambridge, UK: Cambridge University Press.
- [18] Hyland, K. (1998). Persuasion and context: The pragmatics of academic metadiscourse. *Journal of Pragmatics*, 30(4): 437–455.
- [19] Hoey, M. (1979). Signalling in Discourse. No. 6 in *Discourse Analysis Monograph*. Birmingham, UK: University of Birmingham.
- [20] Solov'ev, V. I. (1981). Functional characteristics of the author's abstract of a dissertation and the specifics of writing it. *Scientific and Technical Information Processing* 3: 80–88. English translation of *Nauchno-Tekhnicheskaya Informatsiya*, Seriya 1, Number 6, 1981, 20–24.
- [21] Kircz, J.G..(1998). Modularity: The next form of scientific information presentation? *Journal of Documentation*, 54: 210–235.
- [22] Feltrim, V., Teufel, S., Gracias Nunes, G., and Alusio, S. (2005). Argumentative Zoning applied to Critiquing Novices' Scientific Abstracts. In *Computing Attitude and Affect in Text: Theory and Applications*, James G. Shanahan, Yan Qu, Janyce Wiebe (Eds.) Springer, Dordrecht, The Netherlands, 2005. Pp. 233-245.
- [23] Teufel, S. (2005). Argumentative Zoning for improved citation indexing. In "Computing Attitude and Affect in Text: Theory and Applications" James G. Shanahan, Yan Qu, Janyce Wiebe (Eds.) Springer, Dordrecht, The Netherlands, 2005. Pp 159-170.
- [24] Teufel, S., Siddharthan, A., and Batchelor, C. (2009). Towards Domain-Independent Argumentative Zoning: Evidence from Chemistry and Computational Linguistics. *Proceedings of the 2009 Conference on Empirical*

Methods in Natural Language Processing (EMNLP'09), Suntec, Singapore

- [25] Nordström, Mizuta et al. 2006B. and Ranta, A. (2008). Analyzing Argumentative Structures in Procedural Texts. *Advances in Natural Language Processing, 6th International Conference*, Gothenburg, Sweden, August 25-27, 2008. Lecture Notes in Computer Science, Volume 5221, 2008. Berlin: Springer
- [26] Ait-Mokhtar, S., Chanod, J-P. and Roux, C. (2002). Robustness beyond shallowness: incremental dependency parsing. *Natural Language Engineering*, 8(2/3), pp. 121-144.
- [27] Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*, Chicago: Univ. of Chicago Pr.
- [28] Lisacek, F., Chichester, C., Kaplan, A., and Sandor, A. (2005). *Discovering Paradigm Shift Patterns in Biomedical Abstracts: Application to Neurodegenerative Diseases*. In Proceedings of the 1st International Symposium on Semantic Mining in Biomedicine (SMBM), Hinxtton, Cambridgeshire, UK, 2005.
- [29] Sándor, Á. and Vorndran, A. (2010). The detection of salient messages from social science research papers and its application in document search. *Workshop on Natural Language Processing Tools Applied to Discourse Analysis in Psychology*, Buenos Aires, Argentina, May 10-14. 2010.
- [30] Sándor, Á. and Vorndran, A. (2009). An exploratory system for automatic assistance in peer reviewing research articles in educational sciences. *NLPIR4DL – Workshop on text and citation analysis for scholarly digital libraries. Workshop at ACL-IJCNLP, 2009*. Singapore.
- [31] De Liddo, A., Sándor, Á. and Buckingham Shum, S., Contested Collective Intelligence: Rationale, Technologies, and a Human-Machine Annotation Study. *Computer Supported Cooperative Work*, 21, 4-5, (2012), 417-448. Open Access Eprint: <http://oro.open.ac.uk/31052>
- [32] Synder, C. (2003). *Paper Prototyping: The Fast and Easy Way to Design and Refine User Interfaces*. The Morgan Kaufmann Series in Interactive Technologies, 2003.
- [33] De Liddo, A., Buckingham Shum, S., Quinto, I., Bachler, M. and Cannavacciuolo, L., Discourse-Centric Learning Analytics. In: *Proc. 1st International Conference on Learning Analytics & Knowledge* (Banff, CA, 27 Feb-1 Mar, 2011). ACM: New York. pp.22-33. Open Access Eprint: <http://oro.open.ac.uk/25829>
- [34] Ferguson, R., Buckingham Shum, S. and Deakin Crick, R., EnquiryBlogger – Using widgets to Support Awareness and Reflection in a PLE Setting. In: *Workshop on Awareness and Reflection in PLEs, Personal Learning Environments Conference* (Southampton, 11 July, 2011). Open Access Eprint: <http://oro.open.ac.uk/30598>
- [35] De Liddo, A., Buckingham Shum, S., McAndrew, P., and Farrow, R. (2012). The Open Education Evidence Hub: A Collective Intelligence Tool for Evidence Based Policy. In: *Cambridge 2012: Joint OER12 and OpenCourseWare Consortium Global 2012 Conference* (Cambridge, 16 – 18 April 2012). <http://oro.open.ac.uk/33253>