LAK Explorer – A Fusion of Search Tools

Mike Sharkey
Blue Canary
6185 W. Detroit St.
Chandler, AZ 85226 USA
mike@bluecanarydata.com

Mohammed Ansari
Blue Canary
6185 W. Detroit St.
Chandler, AZ 85226 USA
mohammed@bluecanarydata.com

Andy Nguyen
Blue Canary
6185 W. Detroit St.
Chandler, AZ 85226 USA
andy@bluecanarydata.com

ABSTRACT

The LAK Data Challenge asks the question "What do analytics on learning analytics tell us?" One approach to this challenge is not to answer the question, but to provide a simple, user-focused application that allows any user to easily draw their own conclusion. This was Blue Canary's driver for building the LAK Explorer (http://lakexplorer.bluecanarydata.com). Our team combined multiple tools to create a powerful search application. We extracted topics from the papers, used an autocomplete feature in the search bar, added topics as search result metadata, and provided links to similar papers all as part of the user search experience. The value is in the usability. In the same way that Google presents powerful results via a simple interface, LAK Explorer allows for seamless searching, reading, and comparing of over 400 documents. The application is also instrumented to capture user input (search terms, papers viewed) to provide closed-loop analytics in the future.

Categories and Subject Descriptors

- *Computing methodologies~Natural language processing*
- *Computing methodologies~Algorithms*
- *Information systems~User interfaces*

General Terms

Algorithms, Design, Human Factors

Keywords

Search, natural language processing, similarity, document, vector, elastic search, cosine, autocomplete, corpus

1. INTRODUCTION

The LAK Data Challenge asks the question "What do analytics on learning analytics tell us?" The Blue Canary team tackled this question in 2014 by using topic modeling to describe trends in the LAK Corpus [1]. Topic Modeling was a technique used to distill a large corpus of text into a manageable list of topics. While repeating this approach for LAK15 would theoretically yield new results, it wouldn't do much to advance experimentation in the spirit of the LAK Data Challenge.

Building off of previous LAK entries, the Blue Canary team took a different approach. Instead of analyzing the corpus to look for trends and threads, what if we made the corpus more easily searchable so that the analytics community can browse the corpus for meaning?

This was the core of our approach for 2015. The result is the LAK Explorer (http://lakexplorer.bluecanarydata.com). It's an intuitive search application that allows users to search, browse, and find content in the corpus of papers/articles provided by the LAK Data Challenge. Our goal was to automate the processing so that the LAK Explorer could be applied to any corpus, not just specifically tuned to the LAK data.

1.1 Use of Turbo Topics

As will be explained in this paper, the use of Turbo Topics is a key thread to the Blue Canary team's approach. Blei's research [2] allowed the team to use programmatic techniques to extract n-gram topics from the corpus that end up being a much more user-friendly way to digest corpus content.

1.2 LAK Dataset Incomplete

Blue Canary retrieved the LAK Dataset from the Challenge website (http://lak.linkededucation.org/lak/LAK-DATASET-DUMP.rdf.zip). Upon examination, it appeared as if this dataset did not contain the entirety of the updated 2015 content. Of the 579 content tags (<led:body> and <bibo:content>), 108 were empty. Blue Canary inquired about the gaps but at the time of this project submission, the content was not added to the dataset.

2. The LAK Explorer Components

The LAK Explorer is a fusion of existing tools and techniques for interacting with semantic data. From the simple home screen to the detailed neighborhood of papers, each component was used to add utility to the search process. Figure 1 shows how the different tools were used at different points in the search process.

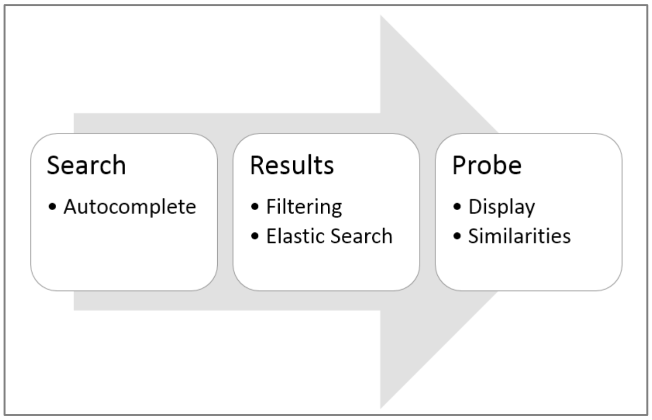
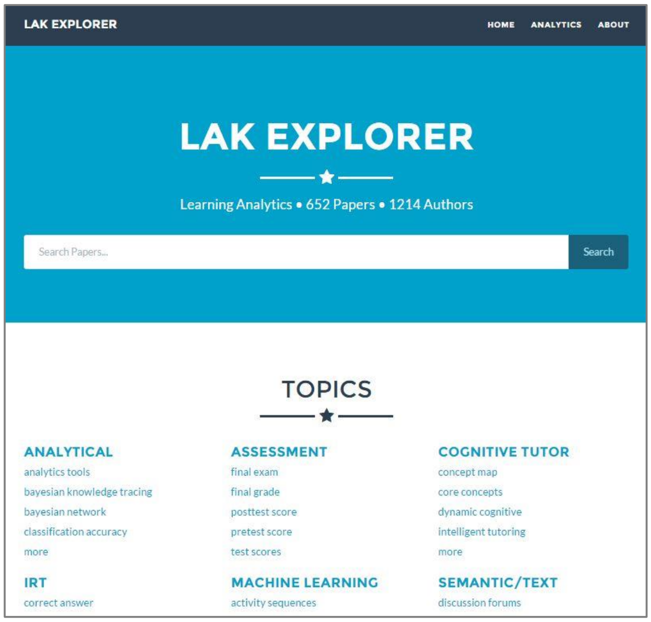


Figure 1. LAK Explorer components

The context for LAK Explorer is similar to that of a search engine. The user comes to the site knowing the universe in which they are searching (corpus of papers) and some idea as to what they want to find out (search terms). However, the user doesn't know exactly what they are looking for. In that way, the presentation of results and related information is vital to improving the utility of the application.

2.1 Home Page

As with any search engine, the usability of the tool starts with the home page. For LAK Explorer, the Blue Canary team was inspired by the simplicity of Google's ubiquitous home page.



LAK EXPLORER

HOME ANALYTICS ABOUT

LAK EXPLORER

Learning Analytics • 652 Papers • 1214 Authors

Search Papers... Search

TOPICS

ANALYTICAL: analytics tools, bayesian knowledge tracing, bayesian network, classification accuracy, more

ASSESSMENT: final exam, final grade, posttest score, pretest score, test scores

COGNITIVE TUTOR: concept map, core concepts, dynamic cognitive, intelligent tutoring, more

IRT: correct answer

MACHINE LEARNING: activity sequences

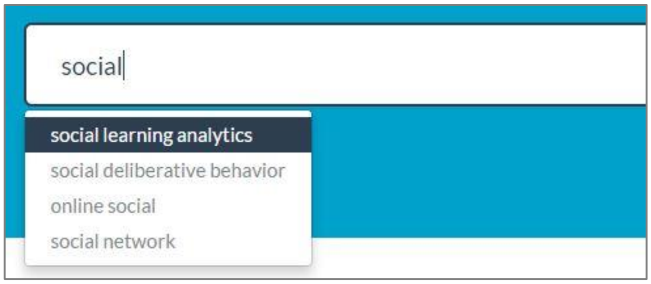
SEMANTIC/TEXT: discussion forums

Figure 2. LAK Explorer home page

The home page is dominated by a large text entry box. The only other significant feature on the page is a listing of topics that were extracted from the corpus of papers.

2.2 Autocomplete

When a user starts entering text into the search box, the first thing they will notice is the use of autocomplete.



social

- social learning analytics
- social deliberative behavior
- online social
- social network

Figure 3. Using autocomplete

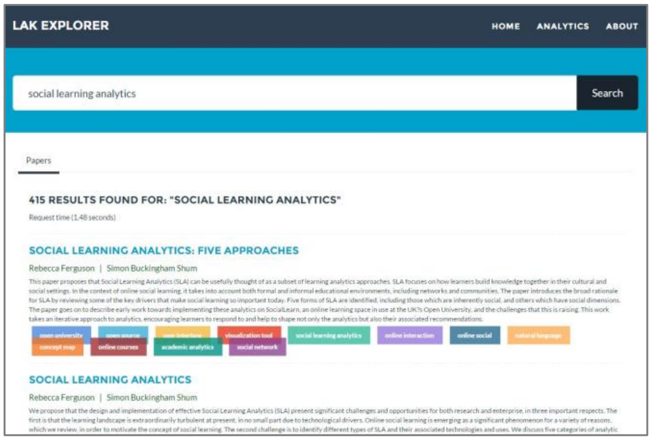
Autocomplete is advantageous since the LAK Explorer deals with a fixed corpus of knowledge. Instead of tying autocomplete to a larger base of content (e.g. dictionary or DBPedia), the team tied it to the topics that were extracted from the corpus using Turbo Topics – the same topics that appear under the search bar. Blue Canary used the Typehead feature from AngularStrap (http://mgcrea.github.io/angular-strap/#/typeaheads#typeaheads) to power this feature.

2.3 Elastic Search

Elastic Search (http://www.elasticsearch.org/) was used to drive the search engine results in LAK Explorer. Blue Canary only used basic features of this tool to drive search results. The text entered in the search box are the inputs to a keyword search algorithm. There are additional features in Elastic Search that could further drive the efficacy of the search and leverage the linked aspect of the LAK data. The search results could be weighted on content found in the content abstract, body, the author(s), and citations.

2.4 Results Filtering

When a user hits the search button, LAK Explorer returns the results similar to the image in Figure 4.



LAK EXPLORER

HOME ANALYTICS ABOUT

social learning analytics Search

Papers

415 RESULTS FOUND FOR: "SOCIAL LEARNING ANALYTICS"

Request time (1.48 seconds)

SOCIAL LEARNING ANALYTICS: FIVE APPROACHES

Rebecca Ferguson | Simon Buckingham Shum

This paper proposes that Social Learning Analytics (SLA) can be usefully thought of as a subset of learning analytics approaches. SLA focuses on how learners build knowledge together in their cultural and social settings. In the context of online social learning, it takes into account both formal and informal educational environments, including networks and communities. The paper introduces the broad rationale for SLA by reviewing some of the key drivers that make social learning so important today. Five forms of SLA are identified, including those which are inherently social, and others which have social dimensions. The paper goes on to describe early work towards implementing these analytics on SocialLearn, an online learning space in use at the UK?s Open University, and the challenges that this is raising. This work takes an iterative approach to analytics, encouraging learners to respond to and help to shape not only the analytics but also their associated recommendations.

SOCIAL LEARNING ANALYTICS

Rebecca Ferguson | Simon Buckingham Shum

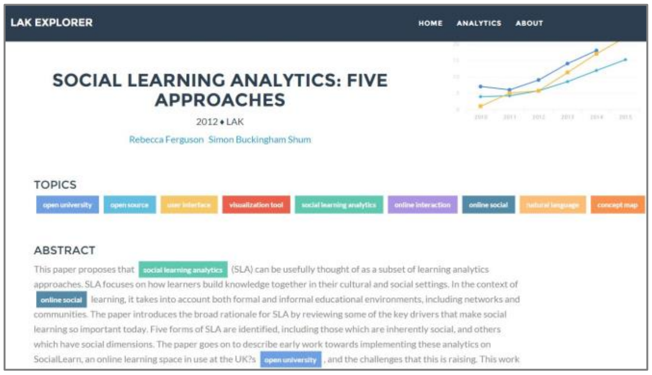
We propose that the design and implementation of effective Social Learning Analytics (SLA) present significant challenges and opportunities for both research and enterprise, in three important respects. The first is that the learning landscape is extraordinarily turbulent at present, in no small part due to technological drivers. Online social learning is emerging as a significant phenomenon for a variety of reasons, which we review, in order to motivate the concept of social learning. The second challenge is to identify different types of SLA and their associated technologies and uses. We discuss five categories of analytic

Figure 4. The search results page

The papers/articles are presented in a fashion similar to an engine like Google Scholar. The title, author(s), and abstract are presented for each result. The most prominent addition is the inclusion of the color-coded topics. LAK Explorer uses the extracted topics as another level of filtering. Upon viewing the results, the user can see what topics are relevant for each paper, click the color-coded topic, and get new search results that are sorted by the frequency of that topic.

2.5 Paper Display

The utility of LAK Explorer is not to just browse search results. Blue Canary wanted to make the tool helpful for actually reading the resulting papers and articles. Therefore, clicking on a paper from the search results gives the display mode shown in Figure 5.



LAK EXPLORER

HOME ANALYTICS ABOUT

SOCIAL LEARNING ANALYTICS: FIVE APPROACHES

2012 • LAK

Rebecca Ferguson Simon Buckingham Shum

TOPICS

ABSTRACT

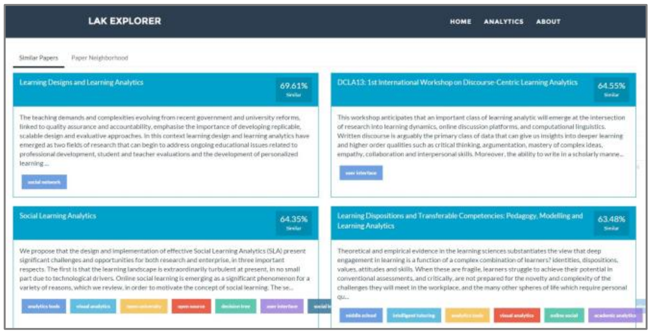
This paper proposes that social learning analytics (SLA) can be usefully thought of as a subset of learning analytics approaches. SLA focuses on how learners build knowledge together in their cultural and social settings. In the context of online social learning, it takes into account both formal and informal educational environments, including networks and communities. The paper introduces the broad rationale for SLA by reviewing some of the key drivers that make social learning so important today. Five forms of SLA are identified, including those which are inherently social, and others which have social dimensions. The paper goes on to describe early work towards implementing these analytics on SocialLearn, an online learning space in use at the UK?s open university , and the challenges that this is raising. This work

Figure 4. The search results page

The paper is displayed in a clear/crisp format for easy online reading. The display is split into three sections. First, the user sees the topics that are most strongly associated with that paper. Then, the abstract is presented followed by the body of the paper. Another key usability point is that the topics are color-coded at the top and the coloring remains intact throughout the body of the paper. This helps draw the reader's attention to topics that may be of particular interest.

2.6 Similar Papers

Perhaps the strongest feature of LAK Explorer is the ability to find similar papers. Keyword and topic searching limits similarity to only papers that share the same frequency of that one term. The similar papers feature uses the entirety of the paper to compare it to other content.



Learning Designs and Learning Analytics 69.61% Similar

The teaching demands and complexities evolving from recent government and university reforms, linked to quality assurance and accountability, emphasise the importance of developing replicable, scalable design and evaluative approaches. In this context learning design and learning analytics have emerged as two fields of research that can begin to address ongoing educational issues related to professional development, student and teacher evaluations and the development of personalized learning...

DCLA13: 1st International Workshop on Discourse-Centric Learning Analytics 64.55% Similar

This workshop anticipates that an important class of learning analytic will emerge at the intersection of research into learning dynamics, online discussion platforms, and computational linguistics. Written discourse is arguably the primary class of data that can give us insights into deeper learning and higher order qualities such as critical thinking, argumentation, mastery of complex ideas, empathy, collaboration and interpersonal skills. Moreover, the ability to write in a scholarly manne...

Social Learning Analytics 64.35% Similar

We propose that the design and implementation of effective Social Learning Analytics (SLA) present significant challenges and opportunities for both research and enterprise, in three important respects. The first is that the learning landscape is extraordinarily turbulent at present, in no small part due to technological drivers. Online social learning is emerging as a significant phenomenon for a variety of reasons, which we review, in order to motivate the concept of social learning. The se...

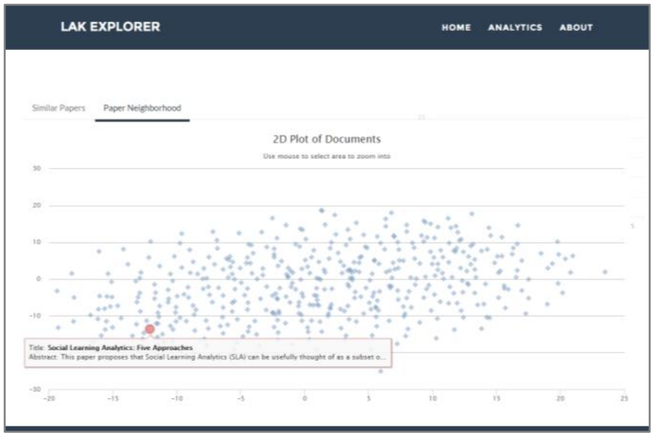
Learning Dispositions and Transferable Competencies: Pedagogy, Modelling and Learning Analytics 63.48% Similar

Theoretical and empirical evidence in the learning sciences substantiates the view that deep engagement in learning is a function of a complex combination of learners? identities, dispositions, values, attitudes and skills. When these are fragile, learners struggle to achieve their potential in conventional assessments, and critically, are not prepared for the novelty and complexity of the challenges they will meet in the workplace, and the many other spheres of life which require personal qu...

Figure 5. Browsing similar papers

Blue Canary used a technique called Doc2Vec [3] to generate similarity scores between two papers. This approach condeses the paper into a single vector, and then a cosine similarity measure is used to compare the vector of one paper to all others in the corpus (http://en.wikipedia.org/wiki/Cosine_similarity). The results (Figure 5.) give a match or score percentage showing papers that are most similar to the one currently being read.

Additionally, LAK Explorer displays this same vector similarity in a neighborhood scatter plot (Figure 6.).



Title: Social Learning Analytics: Five Approaches
Abstract: This paper proposes that Social Learning Analytics (SLA) can be usefully thought of as a subset o...

Figure 6. Visually representing similar papers

The Paper Neighborhood graph takes the Doc2Vec vector data and uses t-SNE (t-Distributed Stochastic Neighbor Embedding) to give all papers two dimensional Cartesian coordinates [4]. The resulting graph shows the current paper (orange dot) in relation to all other papers (blue dots). While the x- and y-coordinates have no real meaning or definition, the spatial representation of each paper allows the viewer to both browse similar papers and to see how "close" or "far" the current paper is from the rest of the corpus.

3. BENEFITS OF LAK EXPLORER

The Blue Canary team wanted to create an application that LAK researchers and practitioners would find valuable. To that end, the team focused on a few key aspects of LAK Explorer to maximize its contribution to the field.

3.1 Visual usability

The Blue Canary product development team gives significant weight to usability. The team might develop an incredibly useful metric, but if the user can't easily interpret that metric, it is useless. A clean layout, color coding, and simple charts all contribute to the usability of LAK Explorer. These features were consciously added in order to improve ease of use.

3.2 Leveraging Turbo Topics

As the team discovered in our LAK14 submission, extracting topics from the corpus turned out to be a simple yet effective way of absorbing the content of papers. We continued this trend for LAK15 by using the Turbo Topics to aid in the initial search and in the meta-tagging of the search results.

3.3 Search Results Plus Similarities

The LAK Explorer was named due to the fact that users will likely not be looking for a singular result. They will look for the results of their search PLUS explore other papers that are similar to their top search result.

The paper similarity tools allow LAK Explorer users to follow this natural path of exploration:

1. I'm interested in papers related to Topic X
2. Searching for Topic X gives me an ordered list of papers
3. I read through Paper 1 that comes up in search results
4. I am also shown Papers 2, 3, etc. that are like Paper 1

4. FUTURE FEATURES

After integrating the previously described components into the LAK Explorer, the team realized that there were additional features we could add to the tool to further increase its value.

4.1 Tracking User Input

The most impactful feature to add is to track user input to the application. LAK Explorer is already instrumented to capture the terms that users search and also the papers that are viewed. The additional feature would be to expose these tracking metrics to the application's front end so that other users can see how the community is interacting with the tool. For example, a simple listing of "most viewed papers' would add more fidelity to other LAK Explorer visitors.

4.2 Linked Data

The linked aspect of the corpus could be further exploited to improve the search process of LAK Explorer. The linked data discerns between abstracts, body, authors, citations, people, and institutions. These parts can be exposed as metadata in the search results (e.g. view more papers by this author) and the data can also be used to drive search efficacy (e.g. weigh hits to the abstract higher than hits to the paper body).

5. SIMILAR INITIATIVES

This is the third year of the LAK Data Challenge and all of the participants continue to stand on the shoulders of previous contributors. Blue Canary acknowledges that previous entrants such as the ones listed in this section have developed toolsets in the same vein as LAK Explorer

5.1 RekLAK

RecLAK [5] was submitted by a team of researchers from PUC Rio in Brazil for LAK14. RecLAK is a recommendation engine that uses the linked nature of the data to recommend other data sources that have feature similarities to the LAK dataset.

5.2 DEKDIV

DEKDIV [6] is an interactive application that allows the user to drill into different aspects of the LAK dataset. The 'Publications' section of DEKDIV was developed in the same spirit as LAK Explorer – allow the user to look at a paper and understand some of the key concepts.

5.3 Visualizing the LAK/EDM Literature

A team of famous LAK researchers submitted a paper to the first LAK Data Challenge in 2013 [7] where, among other things, they clustered the semantic content of the LAK corpus. While using a different technique, this clustering process achieved the same goal of the LAK Explorer's similarity feature set.

6. ACKNOWLEDGMENTS

As with most initiatives at Blue Canary, this was a product of teamwork. LAK Explorer was made possible by a team of players who each brought additive skills to the table. In addition to the named authors, we'd like to thank Satya Mudiam, Faiz Mohammad, Avinash Narasingam, and Kiran Reddy for their contributions.

7. REFERENCES

[1] Sharkey, M., & Ansari, M. (2014). Deconstruct and Reconstruct: Using Topic Modeling on an Analytics Corpus. In LAK Workshops.

[2] Blei, D. M., & Lafferty, J. D. (2009). Visualizing topics with multi-word expressions. arXiv preprint arXiv:0907.1013.

[3] Le, Q. V., & Mikolov, T. (2014). Distributed representations of sentences and documents. arXiv preprint arXiv:1405.4053.

[4] Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. Journal of Machine Learning Research, 9(2579-2605), 85.

[5] Lopes, G. R., Leme, L. A. P. P., Nunes, B. P., & Casanova, M. A. RecLAK: Analysis and Recommendation of Interlinking Datasets.

[6] Hu, Y., McKenzie, G., Yang, J. A., Gao, S., Abdalla, A., & Janowicz, K. (2014). A Linked-Data-Driven Web Portal for Learning Analytics: Data Enrichment, Interactive Visualization, and Knowledge Discovery. In LAK Workshops.

[7] Taibi, D., Sándor, Á., Simsek, D., Buckingham Shum, S., De Liddo, A., & Ferguson, R. (2013). Visualizing the LAK/EDM literature using combined concept and rhetorical sentence extraction.