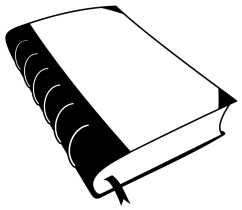

Aberystwyth University

Progress Report

Partridge: An Intelligent Literature Analysis and Recommendation Suite.

Status: Draft

Author:
James Ravenscroft
jrr9@aber.ac.uk
090407039

Supervisor
Amanda Clare Maria Liakata

November 10, 2012

1 Project Description

For scholars carrying out a literature review or researching a specific topic, the vast quantity of research papers available on the internet can often be quite daunting; with the advent of the 'Information Age', choosing which papers to read has become very difficult. This problem is exacerbated when there is a time limit to the research. Partridge aims to aid the reader in finding documents aposite to their interests through the use of Artificial Intelligence. As well as making the research processes more efficient and saving time, the program may also help the scholar to find papers that they may otherwise not have read.

Partridge will use a web interface to allow users to specify particular research topics. This can be done through manual entry of interests and types of report or by allowing the application to collate the papers that the user has read and to learn what is relevant to them. After appropriating this information, Partridge will use learning algorithms to examine and classify the available corpus of literature and to make recommendations as to which articles the scholar would find most helpful.

2 Work to be 'tackled':

Despite the need for a web interface, the main body of the work in this project comes from the backend of the system and the Artificial Intelligence that drives it.

Initial research must be undertaken into Natural Language Parsing techniques and how different features of a text can be used to classify papers. Furthermore, clear identification of the attributes which allow the system to understand the differences between texts is vital.

Once these features are understood, a learning algorithm will be developed to analyse scientific papers and to store information about the differences between them. The system will make use of some third party software (SAPIENTA) to help with this process. Additionally, knowledge from CS36110 Machine Learning will be used in this project.

The first iteration of the system should search for specific criteria corresponding directly to 'features' identified using the aforementioned AI algorithms. Further iterations could be used to build a profile based upon the user's reading history and to make suggestions for further reading.

3 'Project Deliverables'

Although a formal 'XP' methodology will not be adopted since this is an individual project, it will be developed using an agile approach and will be built in an iterative manner. An initial release of the product will be made as soon as possible. Subsequent releases will be made in iterations of two months.

Deliverables include:

- Initial version of the system with restricted functionality; this would use simple search and filtering techniques to find relevant research papers;
- Progress report evaluating the initial product and outlining ongoing development of the Artificial Intelligence underlying the results classifier module;
- Release two of the product, including result classifying behaviour;
- Demonstration of the Artificial Intelligence in the product;
- Following iteration three, the next release of the product will show evolved classification behaviour;
- Final report and release of software.

Initial Bibliography

[1] Awais Athar and Simone Teufel. Detection of implicit citations for sentiment detection. In *Proceedings of the Workshop on Detecting Structure in Scholarly Discourse*, pages 18–26, Jeju Island, Korea, July 2012. Association for Computational Linguistics.

[2] Wiltrud Kessler. Support vector machines and kernel methods in sentiment analysis. `http://www.ims.uni-stuttgart.de/~kesslewd/lehre/sentimentanalysis11s/sentiment_svm_kernels.pdf`, 2011. Accessed: 14/10/2012.

[3] Maria Liakata, Q. Claire, and Larisa N. Soldatova. Semantic annotation of papers: interface & enrichment tool (SAPIENT). In *BioNLP '09: Proceedings of the Workshop on BioNLP*, pages 193–200, Morristown, NJ, USA, 2009. Association for Computational Linguistics.

[4] Maria Liakata, Jee-Hyub H. Kim, Shyamasree Saha, Janna Hastings, and Dietrich Rebholz-Schuhmann. Three Hybrid Classifiers for the Detection of Emotions in Suicide Notes. *Biomedical informatics insights*, 5(Suppl. 1):175–184, 2012.

[5] Naoaki Okazaki. Crfsuite: a fast implementation of conditional random fields (crfs), 2007.

[6] Joseph Townsend, Jim Downing, and Peter Murray-Rust. CHIC - Converting Hamburgers into Cows. In *2009 Fifth IEEE International Conference on e-Science*, pages 337–343. IEEE, December 2009.