Report on the Lucene4IR workshop: Developing Information Retrieval Evaluation Resources using Lucene (L4IR2016)

Leif Azzopardi[1], Yashar Mosfeghi[2], Martin Halvey[1], TBA[3], TBA[3]

[1] University of Strathclyde {*Leif.Azzopardi,Martin.Halvey*}*@strath.ac.uk*

[2] University of Glasgow *Yashar.Mosfeghi@glasgow.ac.uk*

[3] Everyone else to be added {*you*}*@email.com*

September 10, 2016

Abstract

The workshop and hackathon on developing Information Retrieval Evaluation Resources using Lucene (L4IR) was held on the 8th and 9th of September, 2016 at the University of Strathclyde in Glasgow, UK. The event featured three main elements: (i) a series of keynote and invited talks on Lucene in action in industry, in teaching and learning environments, and evaluation forums. (ii) planning, coding and hacking where a number of groups created modules and infrastructure to use Lucene to undertake TREC based evaluations. And (iii) a number of breakout groups discussing challenges, opportunities and problems in bridging the divide between academia and industry, and how we can use Lucene and the resources created in teaching and learning IR evaluation. The event was composed of a mix and blend of academics, experts and students wanting to learn, share and create evaluation resources with the community. The hacking was intense and the discussions lively creating the basis of many useful tools and raising many issues as well as opportunities for the IR community to embrace and utilize the most widely used Open Source IR toolkit for the mutual benefit of academics, students, researchers, developers and practitioners.

1   Introduction

Lucene and its expansions, Solr and ElasticSearch, represent the major open source Information Retrieval toolkits used in Industry. However, there is a lack of coherent and coordinated documentation that explains from an experimentalist?s point of view how to use Lucene to undertake and perform Information Retrieval Research and Evaluation. In particularly, how to undertake and perform TREC based evaluations using Lucene.

The objective of this event is to bring together researchers and developers to create a set of evaluation resources showing how to use Lucene to perform typical IR operations (i.e. indexing, retrieval, etc.) as well as how to extend, modify and work with Lucene to extract typical statistics, implement typical retrieval models, and to evaluate various TREC tasks. Over the course of days, we essentially plan to have a hackathon to build these resources,

share knowledge about the platform, and to create a road map for future development. The tools and resources built during the workshop/hackathon will be uploaded to GitHub and shared with the community.

2   Keynotes and Invited Talks

Keynote Talk: Lucene in Industry - Charlie Hull (Flaxsearch)
Flax have been building search applications using open source search software since 2001. Charlie will talk about the rise of Lucene-based search engines and how Flax have used them in a number of different real-world projects. He'll also discuss some ideas about improving the links between academia and industry.

Using Lucene for Teaching and Learning IR: The University of Granada case of study ? Prof. Juan Manual Fernandez Luna (University of Granda)
In this talk, we shall describe how the University of Granada is supporting teaching and learning Information Retrieval (TLIR) discipline across different courses in the Computer Science studies (degree and masters), and how this is done by means of the Lucene API. Later we shall present our thoughts about how Lucene could be used inTLIR context and present some proposals for improving the Lucene experience, both for students and lecturers.

Evaluation and Reproducible Experiments - Sauparna "Rup" Palchowdhury (NIST)
Having seen students and practitioners in the IR community grapple with abstruse documentation that accompany search systems, I want to direct some attention to Lucene's internals and demonstrate how to do IR experiments using it. Ensuring that a search system you have built works 'correctly' entails evaluating its output on test-collections like those from TREC. In this way evaluation finds a purpose in my exposition. The goal is to help bring Lucene to the IR community and prevent its usage as a black box, which misleads students because they learn little from their results. On top of that it makes it hard to reproduce experiments reported in papers. The talk will also go over Lucene's scoring by showing how a TFxIDF term-weighting scheme is implemented. I intend to initiate a discussion on implementing models of similarity and accept feedback to validate the implementation I describe.

To this end, I have built tools and written notes to help the experimenter organize her work: http://kak.tx0.org/IR/. Here is a glorified Python script called TRECBOX that runs other search engines, notes on Terrier (TTR) and Lucene's (LTR) internals. LTR is a 'mod' that works on TREC data to help in quickly getting a system up and running; tests prove it to behave correctly on TREC test-collections.

Deep Dive into the Lucene Query/Weight/Scorer Java Classes - Jake Mannix (Lucidworks)
To provide some overview of how to do scoring modifications to Lucene-based systems, we'll get down and dirty with the BooleanQuery class and its cousins, seeing where the Lucene API allows for modification of scoring with pluggable Similarity metrics and even deep inner-loop ML-trained ranking models - *if* you're willing to do a little work.

Learning to Rank with Solr - Diego Ceccarelli (Bloomberg)
Learning To Rank is a technique that allows you to apply a machine learning model in the construction of a ranking model for information retrieval systems. In fact, it is well known that sophisticated models can make more nuanced ranking decisions than a traditional

ranking function. At Bloomberg we have integrated a learning to rank component directly into Solr, enabling others to easily build their own Learning To Rank systems and access the rich matching features readily available in Solr. In this talk we will present the key concepts of Learning to Rank, how to evaluate the quality of the search in a production service, and then how the Solr plugin works.

3 Discussion

Challenges and Opportunities
    Teaching and Learning
    Feedback

4 Resources

5 Summary

6 Acknowledgments