Mining Academic Expertise from Funded Research (Expert Search System)

Peeranat Fupongsiripan 2056647

February 4, 2014

Abstract

This project is concerned with development of mining academic expertise in Scottish universities from funded research. This system can be used to extract, analyse and find experts in particular areas in Scottish universities. http://experts.sicsa.ac.uk/ is an existing academic search engine that assists in identifying the relevant experts within Scottish Universities, based on their recent publication output. The aim of this project is to develop mining tools for the data, and research ways to integrate it with existing deployed academic search engines to obtain the most effective search results. Most of the project's aims and requirements were satisfied. However, various difficulties were encountered in the late stages of the development life-cycle which ultimately led to a reduction of the system's scope.

Acknowledgements

Thanks to Prof. Craig Macdonald for his advice and supervision throughout this project.

Contents

1 Introduction

1.1 Definition of Mining Academic Expertise from Funded Research (Expert Search System)

In an expert search task, the users' need is to identify people who have relevant expertise to a topic of interest. An expert search system is an Information Retrieval(IR) system that can aid users with their "expertise need". Moreover, it predicts and ranks the expertise of a set of candidate persons with respect to the users' query. Mining academic Expertise from funded research is to develop tools for funded research data which can extract, analyse and retrieve those data.

1.2 Why Implement one?

With the advent of the vast pools of information and documents in large enterprise organisations, collaborative users regularly have the need to find not only documents, but also people with whom they share common interests, or who have specific knowledge in a required area. The system makes use of textual evidence of expertise to rank candidates.

1.3 Aims

http://experts.sicsa.ac.uk/ is an existing academic search engine that assists in identifying the relevant experts within Scottish Universities, based on their recent publication output. The aim of this project is to develop mining tools for the data (funded research), and research ways to integrate it with existing deployed academic search engines to obtain the most effective search results. Learning to Rank Algorithm for Information Retrieval (IR) is also used in this project in order to enhance the effectiveness of the system. Funded projects from Grant on the Web are obtained from - http://gow.epsrc.ac.uk/ and Research Councils UK - http://gtr.rcuk.ac.uk/. Academic funded projects and publications are used as expertise evidence to assist in identifying the relevant experts with respect to a user query.

1.4 Prerequisites

Where possible all technical concepts used within this paper will be clearly defined. However, to fully understand the remainder of this report, the reader should have at least some knowledge of the following concepts:

- Object-Oriented programming concepts (preferably in Java)
- Concepts of Information Retrieval (IR)
- Basics of Learning to Rank Algorithms for IR (preferably AdaRank)

2 Requirements Specification

These requirements were decided on during the early stages of the project. The requirements have been split into several sections depending on their necessity

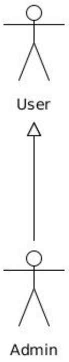


Figure 1: Actors

2.1 Actors

Figure 1 illustrates the relationships between actor roles in the system. A short summary of the actors is given below:

User represents an actor in the system who is able to search for experts with respect to a query. The search can be done via SICSA look or university template look.

Admin inherits from user and is able to modify look and feel, update academics and view queries infromation.

2.2 Use Cases

The core use cases for the system are:

- Administration

  Login

  Logout

  View Status

  View Queries Frequency

  Change the way system updates academics

  Generate a custom look for university

  View publications details and visits statistics

- Common utility activities

  Search for experts with respect to a query

  View Academic Profile

  View relevant publications with respect to a query

  View relevant funded projects with respect to a query

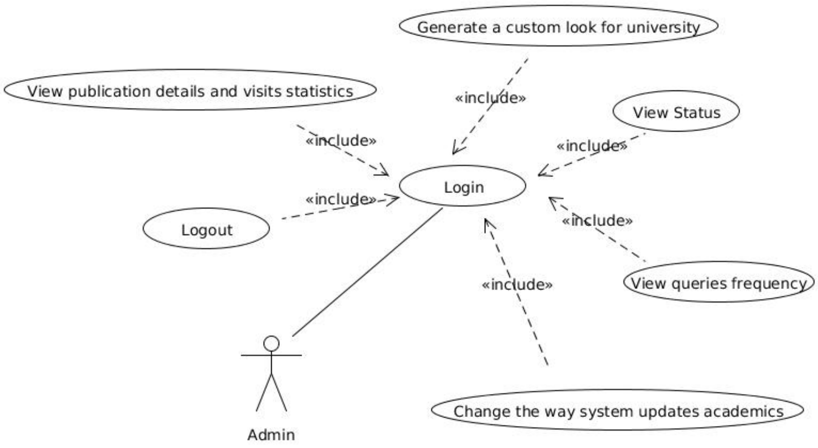
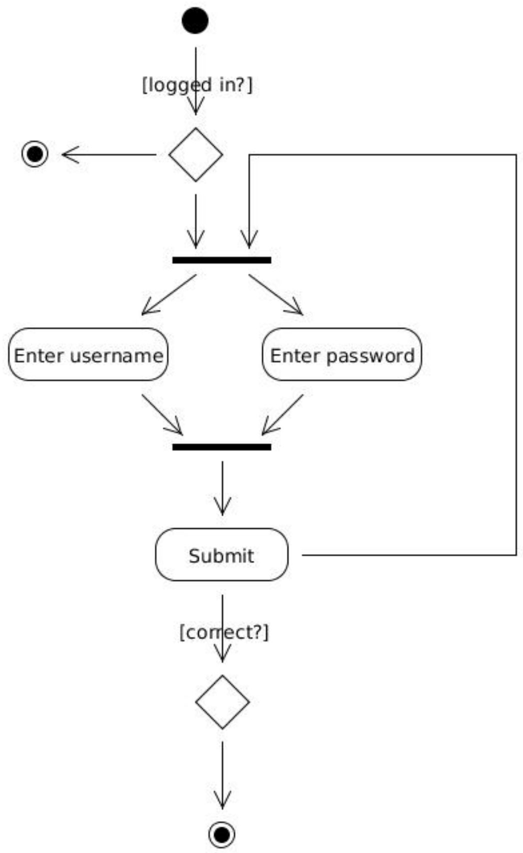

Generate a custom look for university
View publication details and visits statistics
«include»
View Status
«include»
«include»
Login
«include»
Logout
«include»
«include»
View queries frequency
Admin
Change the way system updates academics

Figure 2: Use Case Diagram

2.2.1 Administration

Figure 2 illustrates the use case diagram of the system.

Use Case	Login
Description	 [logged in?] Enter username Enter password Submit [correct?]
Rationale	The login use case allows administrator to access to the system. The usename and password are provided by SICSA staff
Priority	Must Have
Actors	- Admin
Conditions	**post** The user is logged in if correct credentials are provided

Use Case	Logout
Description	The logout use case changes a user's status to logged out. Logout is invoked by the user
Rationale	The logout use case allows a user to leave to system to prevent unauthorised use from an unattended terminal
Priority	Must Have
Actors	• Admin
Includes	Login
Conditions	• **post** The user is logged in

Use Case	View Status
Description	A user's status (institution, email) is displayed
Rationale	User needs to be able to view their status in the system
Priority	Should Have
Actors	• Admin
Includes	Login
Conditions	• **pre** The user is logged in

Use Case	View Queries Frequency
Description	The submitted queries Frequency associated to user's institution is displayed
Rationale	User needs to be able to view queries frequency in order to determine the popularity of particular areas
Priority	Should Have
Actors	• Admin
Includes	Login
Conditions	• **pre** The user is logged in

Use Case	Change the way system updates academics
Description	Initially, the system scrapes the academics from user's webpage. Some users may prefer XML feeds from each university. There are currently 2 options: • an XML feed hosted on user's website and accessed directly by the system • manual uploading of XML file to the system
Rationale	Allows user to upload academics in different ways
Priority	Should Have
Actors	• Admin
Includes	Login
Conditions	• **pre** The user is logged in • **post** Change updated