Dublin City University

School of Computing

CA4009: Search Technologies

Designing Search Solutions

Gareth Jones

October 2019

Introduction

The increase in digital information archives is creating opportunities and needs for search technologies in many settings.

The specification of an effective search system for a particular application requires an understanding of the task or tasks that the user of the system will need to be able to perform.

Satisfying these user requirements will generally pose challenges for search technologies,

Meeting these challenges requires selection of the most appropriate available search technologies to address them.

Search System Specification

Similar to other interactive software systems, developing effective search applications requires analysis of the users, the tasks of the users, the setting within which the system is to be used, and the platform to be used.

However, search applications also require an understanding of the data to be indexed, and the challenges and opportunities it presents for search.

Once these have been understood, possible search technologies (algorithms, file structures, etc.) can be selected.

It is often not theoretically or even intuitively clear which technologies represent the best choice for use in a particular search system.

Search System Design and Implementation

Even for individual technologies, there are often parameters which affect the behaviour of individual algorithms which need to be selected in order to optimise their effectiveness in specific settings.

Appropriate choice of these parameters can then be vital to achieving the most effective setup of a search system.

In such cases the only way to choose the best solution is to evaluate the component or overall system for its effectiveness for this search task, and where appropriate to investigate the impact of parameter settings.

The performance of search systems incorporating alternative components, technologies and/or parameters can then be compared, and the most suitable ones selected.

Search System Design and Implementation

The chosen technologies may have the best absolute retrieval effectiveness or there may be some trade-off, e.g. over the computational cost of the method vs its retrieval effectiveness.

This requires a choice (or design) of suitable evaluation metric(s) - where suitable indicates that these metric(s) provide some usefuyl or meaningful measure of the system's effectiveness in the context of the task application for which it will be used.

If not already available. a suitable information retrieval test collection must also be defined and constructed.

This test collection is then indexed into the search system and the values of the evaluation metrics calculated for the queries contained in the test collection.

New Technologies for Search Solutions

In some situations the requirements of the desired search system cannot be satisfied using existing components or technologies.

In such situations, the system designer must create a novel technology solution.

Novel components are often based on existing technologies or “borrow” idea or methods from other systems, e.g. other natural language technologies or image or video processing technologies.

New Technologies for Search Solutions

These new technologies typically form a component of a search system which is otherwise composed of existing components.

As before, if may not be clear whether a new idea will be effective or more effective than another proposed solution.

In these cases, the proposed solution(s) must be evaluated and compared - and if not effective must be revised or replaced.

Example Problems

Problem 1: Multi-source Search for Journalists: New search system for news journalists to search multiple news and current affairs sources.

Typical questions to be addressed for a new search system:

- What are the search requirements?
- Who are the users?
- What data will be searched?
- What technologies are available?
  Are there required technologies which do not current exist and need to be developed?
- How can the effectiveness of the new system or its components be evaluated?

Examples Problems

Problem 2: Desktop Search: Integrated search of all content stored on a desktop or laptop computer.

Problem 3: Culinary Recipe Search: Searching for suitable recipes using the ingredients available in your kitchen.

Problem 4: Search for Learning: Incorporating facilities to support learning about a topic while searching on the topic.

In all cases, evaluation tasks must be selected or developed, and a suitable test collection developed.