

Dialogue-based Search Solution

Masters Thesis proposal (in cooperation with Findwise & Talkamatic)

1. Background

In an Enterprise environment a search system often crawls and indexes a large number of different data sources – databases, Content Management Systems, external web pages, file shares with different types of documents, etc. Each of the data sources or sub sources may have a primary target group – e.g. Sales, Engineers, Marketing, Doctors, Nurses, etc. all dependent on the type of organization.

The purpose of the (unified) Search system is to serve as a platform (a single entry point) to satisfy the information need for all the different groups in an organization. However, given that search queries are often short (~2.2 words) and ambiguous, and the users have different background, the system employs a number of techniques for filtering of and drill down into the search results. One such technique is facets, e.g. a filtering based on data source, additional keywords, dates, time, etc.

On the other hand, there are at least two types of users (behaviour) – those that know exactly what they look for and how to find it, and use the search in stead of menu clicks; and those who do not know exactly what they look for, nor where the potential information may be found. We can consider these two groups as the two extremes in much fine-grained scale.

We would like to concentrate on the second group of users, who often engage in some sort of dialogue with the search system. Such users may interact in several ways with the system during a search session – they may rewrite and expand they original query, they may filter it by facets, click on some documents until they finally discover or not what they were looking for.

2. Dialogue Systems

Spoken dialogue systems are computer systems which use speech as their primary output and input channels. Dialogue systems are primarily used in situations where the visual and tactile channels are not available, for instance while driving, but also to replace human operators for instance in call centers. Recently, spoken dialogue systems have become more widespread with the arrival of Apple's Siri

and Google's Voice Actions, even outside of the traditional areas of use. As speech and voice has the potential of transmitting large quantities of information very fast compared to traditional GUI interaction, this is a development which is likely to continue.

A spoken dialogue system typically consists of a dialogue manager, an Automatic Speech Recogniser, a Text-to-speech engine, modules for interpretation and generation of utterances and finally some kind of application logic.

Voice search is a term which has emerged the last years. The user speaks a search query, and the system responds by returning a hit list, much like an ordinary Google search. If the hitlist doesn't contain the desired hit (document, music file, web site etc.) the user needs to do a new voice search with a modified utterance.

The idea of this project is to replace voice search by dialogue-based search, where the user and the system engage in a dialogue over the search results in order to refine the search query.

3. Dialogue-based Search – case study

The task of the Masters thesis is to explore the possibilities of using Dialogue-Systems/Dialogue acts in order to satisfy the information needs of certain groups of users in a Search system. The target group consists of several types of users:

- Users who submit very broad and ambiguous search queries (e.g. "Greece", "food", "pm")
- Users who do not employ the tools provided by the Search system such as facets (e.g. queries such as "pm pdf")
- Users with exploratory queries (e.g. "Abba first album")

4. Document format – details

Before documents are being sent for indexing in the Search System, they have been augmented with META-data. The metadata allows us to do a number of things:

- Advanced queries
- Filtering
- Sorting
- Faceting
- Ranking

The format of the indexed document could look like:

```
<doc>
  <field name="id">6H500F0</field>
  <field name="name">Maxtor DiamondMax 11 - hard drive - 500 GB - SATA-300</field>
  <field name="manufacturer">Maxtor Corp.</field>
  <field name="category">electronics</field>
  <field name="category">hard drive</field>
  <field name="features">SATA 3.0Gb/s, NCQ</field>
  <field name="features">8.5ms seek</field>
  <field name="features">16MB cache</field>
  <field name="price">350</field>
  <field name="popularity">6</field>
  <field name="inStock">>true</field>
  <field name="manufacturedate_dt">2006-02-13T15:26:37Z</field>
</doc>
```

```
<doc>
  <field name="id">1</field>
  <field name="title">London</field>
  <field name="body">London is the capital of UK. London has 7.8 million
inhabitants</field>
  <field name="places">London</field>
  <field name="date">2012-11-30</field>
  <field name="author">John Pear</field>
  <field name="author_email">john@pear.com</field>
  <field name="author_phone">+44 123 456 789</field>
</doc>
```