# Natural Language Access to Yellow Pages

**A De Roeck, U Kruschwitz, P Scott, S Steel, R Turner and N Webb**
Department of Computer Science, University of Essex, Colchester, UK
{udo,webbnw}@essex.ac.uk

## Introduction

The YPA project (De Roeck et al. 1998) is building a system[1] to make the information in classified directories more accessible.

There are two reasons for doing this:

- Directories contain much useful but hard-to-access information, especially the free text in semi-display advertisements; finding an advertisement requires prior knowledge of the indexing employed by the specific directory;
- More generally, the project is a demonstrator for exploitation of semi-structured data - data that is less systematic than database entries or logical clauses, but more systematic than free text because it has been marked up, for display or some other purpose.

This task requires a narrowing of the gap between user requirements and the indices offered by the domain. To achieve this, simplified approximations of natural language processing are used to soften the interface, which retain some structure of the input query allowing more precise access to specific advertiser information. From the input data, knowledge based indices are built, which capture more accurately the information stored in the directory than standard information retrieval techniques.

For those queries that the system is unable to answer in a single question-answering cycle, there is a dialogue management system that sits between the robust natural language *Frontend* and the knowledge based indices contained in the *Backend*. This *Dialogue Manager* contains knowledge both of the domain and also some cheap common sense world knowledge, which allows the *Dialogue Manager* to reason about possible variations to the user input, by using synonym and hyponym relations, for example.

The ability of the *Frontend* to retain the input query structure enables greater flexibility during the construction of a query for the *Backend*.

BT's *Yellow Pages*[2] provides an example of a classified database with which this work would be useful, and forms the basis of the on-line demonstrator of the YPA.

## System Overview

A conversation cycle with the YPA can be roughly described as follows. A user utterance (typed in via the *Graphical User Interface*) is sent to the *Dialogue Manager*. The *Dialogue Manager* keeps track of the current stage in the dialogue and controls the use of several sub-modules. Before handing back control (together with the relevant data) to the *Toplevel*, the input is first sent to the *Natural Language Frontend* which returns a so-called *slot-and-filler query*.

The *Dialogue Manager* then consults the *Query Construction Component*, passing it the result of the parsing process (possibly modified depending on the *dialogue history* etc). The purpose of the *Query Construction Component* is to transform the input into a *database query* (making use of the *Backend* and possibly the *World Model*), to query the *Backend* and to return the retrieved addresses (and some database information) to the *Dialogue Manager*.

Finally the *Dialogue Manager* hands back control to the *Toplevel* which for example displays the retrieved addresses. It could also put questions to the user, which were passed to it by the *Dialogue Manager*, if the database access was not successful (i.e. did not result in a set of addresses). At this stage the cycle starts again.

[1] This work has been funded by a contract from the ISR group at BT Laboratories, Martlesham Heath, UK.

[2] Yellow Pages® and Talking Pages® are registered trade marks of British Telecommunications plc in the United Kingdom

Accepted as a short presentation at the 8th International Conference on Human-Computer Interaction (HCI99), Munich, August 1999.

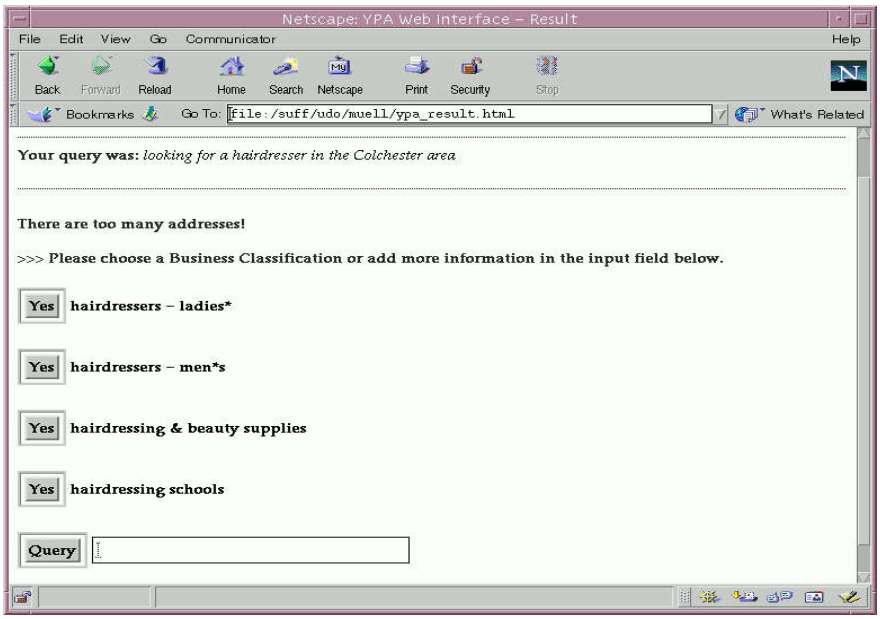


Figure 1 – Screenshot of the YPA

## Frontend

The methods used to extract *slot-and-filler queries* from natural language input are more fully described in (Webb et al. 1999). They consist mainly of a shallow, bottom up parser and a slot-filling mechanism. This mechanism tries to preserve some of the parse structure in domain-dependent slots, which represent the most important roles in the interaction. These are goods, location and transaction in the YP domain.

The *Dialogue Manager* (*DM*) is outlined in (De Roeck et al. 1998), and is currently a transition-based approach. The *DM* controls all interactions between the *Backend* database, the user and the various knowledge sources employed by the system.

## Backend

The *Backend* construction process takes the raw data (the YP printing tape) and creates a database that retains as much information and relations as possible for the online enquiry system. The input data is semi-structured in a sense that a record structure for the addresses and headings does exist but the internal structure of these entries is not formally defined. Usually this consists of partial English sentences, address information, telephone patterns, among others. This process was described in detail in (De Roeck et al. 1998). (Kruschwitz et al. 1999) report on recent experiences in the extraction of semi-structured data.

## Future Work

We continue work on the YPA system. We now have access to richer data sources and components such as the *Dialogue Manager* are being re-implemented using new mechanisms, as a new *Backend* database allows refined query construction and a more user-friendly dialogue.

A major task for the future is a deep evaluation of the system, involving a user trial, once we have developed a framework for this.

## References

De Roeck, A., Kruschwitz, U., Neal, P., Scott, P., Steel, S., Turner, R. & Webb, N. (1998). YPA – an intelligent directory enquiry assistant. *BT Technology Journal*, 16(3), 145–155.

Kruschwitz, U., De Roeck, A., Scott, P., Steel, S., Turner, R. & Webb, N. (1999). Extracting Semi-Structured Data – Lessons Learnt. *Proceedings of the AAAI Fall Symposium on Using Layout for the Generation, Understanding or Retrieval of Documents*. Massachusetts, November 1999. *Forthcoming.*

Webb, N., De Roeck, A., Kruschwitz, U., Scott, P., Steel, S., & Turner, R. (1999). Natural Language Engineering: Slot-Filling in the YPA. *Proceedings of the Workshop on Natural Language Interfaces, Dialogue and Partner Modeling*. Bonn, September 1999. *Forthcoming.*