MASTERS in COMPUTER SPEECH, TEXT AND INTERNET TECHNOLOGY

NAME: Nicolaas Matthijs

College: Queens' College

CRSID: nm417

Complete All Sections

1.   PROJECT CHOICE IN ORDER OF PREFERENCE:

i)   Project Title: Improving Click-Through Rates using Keyword Search Terms
        Supervisor: Stephen Clark

ii)   Project Title: Unsupervised graded word sense disambiguation
        Supervisor: Diarmuid Ó Séaghdha

iii)   Project Title: Speaker adaptation by voice conversion for HMM-based speech synthesis
        Supervisor: Heiga Zen (Toshiba) / Mark Gales (CUED)

2.   PROJECT SUGGESTION:

Project Title:	Personalized Search
External Supervisor:	Dr. Filip Radlinski (Microsoft Research)
Internal Supervisor:	Dr. Stephen Clark
Project Content:	Different people look for different things with the same queries. Classic examples include queries like jaguar, columbia, flash, mercury. This project will involve creating a personalized search system.

The first part of the project involves generating an interest profile for the user. This will be based on a variety of data sources, like browser history, e-mail, files on computer, previous search queries, Facebook profile, … In order to capture all of this data, additional software (e.g. browser plug-in) will be written.

The second part of the project will involve using this user profile to predict which Wikipedia disambiguation is the most interesting to a user or to re-rank the output of a search engine. The results will then be evaluated in the third part of the project. I will also assess which of the used data sources describes the user best.

3.   **COMMENTS:** (e.g. any special arrangements made with potential supervisor)

I have agreed to meet weekly with my external supervisor.

4.   **SPECIAL RESOURCES:** (e.g. any special software, extra disk space, licences)

Please note that these are to be discussed and agreed with the potential supervisor and, for projects being carried out with Computer Laboratory machines/filespace, Mr Chris Hadley before starting the project. Special resources required for speech-based projects should be coordinated with the supervisor in the CUED concerned.

*Special resources - extra filespace etc.

N/A

*Experiments on Human Subjects

If your project involves experiments on human subjects, then please tick below and attach a short justification.
[X] My project involves experiments on human subjects (justification for experiments attached)

Because the project will try to personalize Wikipedia disambiguation and search engine re-ranking according to a specific user's interests, evaluation will need to happen on the people that have provided data for the project. One scenario is to bring people in to the Computer Labs and have them manually evaluate the Wikipedia disambiguations chosen for them. An alternative scenario is to ask volunteers to run software on their computer that allows them to provide data and can produce personalized results in real life.

*Using your own computer

If applicable, tick below and complete both sections A and B and sign where indicated.

[X] I plan to use my own computer

A. The computer I'll be using for my project has the following configuration (e.g. 3 GHz, 4 Gb RAM, 300 Gb Disk, Linux OS)

MacBook Pro 2.66 GHz, 4 Gb RAM, 250 Gb Disk, Mac OSX Snow Leopard

B. I accept full responsibility for this machine and I have made contingency plans to protect myself against hardware and/or software failure.

_____

*Other resources:

This section is for anything that is not covered above, including specialist hardware or software or unmanaged machines belonging to a research group. For each item we need to know who has accepted responsibility for providing it (this may be yourself).

Item description   Signature of provider   Printed name
N/A

Declaration:

To the best of my knowledge and belief, I have declared all of the resources I will need for completion of my project.

_____