

Natural Language Processing & Applications

Continuous Assessment (20% of Module Mark), 2007/08

You are employed in the research division of a company whose main activity is the provision of IT services to other companies. Some 60% of the company's revenue comes from the provision of e-commerce services. The task of your division is to produce reports on promising new areas, perhaps still only at the research and development stage, so that the company can decide on its strategy: which existing areas to continue to invest in, whether to develop new areas, etc.

Your Division Head has recently become interested in the idea of employing NLP in e-commerce. She was inspired by reports of the Beijing Virtual Economy District (also called the Beijing Cyber Recreation District), where it is suggested that avatars could be used to allow customers from all over the world to order Chinese-made goods. However, she believes that NLP could have wider application in e-commerce generally and so may be worth investing in.

She intends to produce a report covering not only the specific NLP technology which would be needed, but also the background to the speech recognition, machine translation and speech synthesis. Along with other colleagues, you have been asked to prepare **part** of the report, which she will edit together with the addition of overall conclusions and recommendations.

Choose **one** of the two areas speech synthesis/recognition or machine translation. Write the relevant part of the report. The absolute maximum length is 2,000 words (excluding any figures, references, etc.). Length is not of itself a virtue; a well-researched and well-argued but brief report is preferable to a 'woolly' or 'padded' one. **Referencing must be done to academic standards.**

By **prior agreement with me**, two people can prepare a joint report. In this case the absolute maximum length is 3,500 words. Equal marks will be awarded so it will be your responsibility to ensure equal effort.

The report should contain a statement of its word length.

I reserve the right to ask you to supply your report as a Word document or text file, so the report must be word-processed and you should make sure you keep the file.

Your report should be handed in to the School of Computer Science Reception by 12 noon on **17 January 2008**. Penalty for lateness: 5% of the total mark per working day or part thereof.

My target is to return the marked work within 4 weeks of the hand-in date. However, I am teaching part of a module I have never taught before in the first half of Term 2, and the School has an accreditation visit by the BCS which I am running, so I'm quite likely to miss this target.

Marking Notes

In order to complete the continuous assessment well, you will need to do additional reading and research beyond the lectures. Searching the web combined with some library work will provide many sources.

I'm not looking for any fixed answer, so I'll mark by impression, putting myself in the position of your Division Head. The marking scheme will be based on five components, initially with equal weights:

- scenario: appropriate use of the context/scenario described above
- research: use of relevant background research
- understanding: demonstrated understanding of any necessary theory and methods (but without excessive detail)
- outcome: deployment of the evidence to form a coherent (but **partial**) set of recommendations
- style: an appropriate style for an internal company report, with excellent use of English.

I won't necessarily just add the component marks, since different answers may emphasize different components.

For an A grade (70% and over), I would expect to see all five of these components present to a high standard.

For a B grade (60% and over), there might be less evidence of relevant background reading but still a good knowledge of any necessary theory and methods, a coherent argument and clear recommendations, and a good style.

For a C grade (50% and over), I'd need to see some evidence of relevant material outside the lectures, a reasonable understanding of any necessary theory and methods, coupled with perhaps a less coherent analysis or set of recommendations.

For a D grade (40% and over), either evidence of relevant material but not well deployed or a good report based on marginal research.

A Fail would be indicated by little or no use of relevant material and/or incoherent arguments or recommendations.

The averages for all students for the five components in last year's work are worth noting:

Scenario	53%
Research	69%
Understanding	68%
Outcome	51%
Style	70%
<hr/> Overall	<hr/> 63%

The two weak areas are clear; both are connected with the industrial/commercial context, in particular the need to reach evidenced recommendations relating to future investment, rather than produce general evaluations of NLP techniques.