
MSc CS Project Form

Consult the [MSc CS programme intranet page](#) for information about the project. The assessment criteria which examiners apply are reproduced in Section 4 at the end of this form.

1. Brief proposal

After agreeing on a title and brief outline of a project with a supervisor, the student should complete part 1 of this form. There are default values for the weights of the marking aspects for both the proposal and the report. They can be altered in agreement with the supervisor for both the extended proposal (Section 2a and 2b) and the final project (Sections 3a, 3b and 3c). The weights must stay within the ranges given in brackets and must add up to 100. Students should upload the completed form to [Moodle](#). See the programme intranet for more details.

Student details

Student ID	13006376
Name	Matthew Ralph
Programme	MSc Computer Science (Part-Time)
Supervisor	Dell Zhang
Email	mralph02@mail.bbk.ac.uk

Project details

Title: Hansard NER - character-based Named Entity Recognition for the Hansard using Birkbeck SAMTLA

Proposal outline (approximately 200 words): To download, pre-process and load the Hansard record of the House of Commons, into a copy of the Birkbeck/British Library Search and Mining Tool for Language Archives (SAMTLA) system.

Then, to augment the text mining functionality of SAMTLA with character-based Named Entity Recognition (NER) (using CharNER implementation of Kuru, Can and Yuret from https://github.com/0xnurl/keras_character_based_ner), thus improving SAMTLA's current NER, which is based on a static gazetteer. As a new Hansard report is created on every day that parliament sits, a more dynamic form of Named Entity recognition is required, as new terms and references will be introduced throughout the life of this living dataset (cf Heaps' law).

Lastly, to augment the front-end of the SAMTLA interface (Javascript) to highlight these named entities, and the type of entity that they are. A possible extension is to display percentage confidence in entity type, as calculated by CharNER.

The system will have several discrete components:

1. System for downloading Hansard using the TheyWorkForYou API (<https://www.theyworkforyou.com/api>), which can be scaled up to download multiple days' Commons debates in a clean format.
2. Pre-processor to split Hansard into characters and perform any other required pre-processing
3. Scripts to load a simple, character-based representation of the Hansard into CharNER for tagging
4. Scripts to load the dataset into SAMTLA's k-truncated suffix-tree data format, complete with its Named Entity information
5. Frontend work to display the Named Entities, possibly including percentage confidence in a Named Entity belonging to a particular class, e.g. Person, Organisation.

- Work plan:**
1. Literature review on character-based NER approaches done
 2. Work review of the open-sourced codebases implementing character-based NER approaches
 3. Develop scripting to download Hansard Commons reports from TWFY(They Work For You) API, in XML format, and perform any necessary pre-processing.
 4. Review existing SAMTLA code-base, and existing gazetteer NER implementation.
 5. Set up a SAMTLA copy containing a 'toy' Hansard dataset
 6. Run a copy of CharNER or similar, training it with CoNLL dataset (e.g. <https://www.clips.uantwerpen.be/conll2003/ner.tgz>) or similar
 7. Provide integration with Samtla system for NER determinations

A full plan is given in the proposal document.

Departmental equipment/software required (contact the Systems Group for advice):

Access to Samtla system to load data and integrate Named Entity integration

Weights agreed with supervisor: yes

All ethical issues are routine (agreed with supervisor): YES (delete as appropriate)

If no, consult [http://www.bbk.ac.uk/committees/researchintegrity/
GuidelinesonResearchwithEthicalImplications.pdf](http://www.bbk.ac.uk/committees/researchintegrity/GuidelinesonResearchwithEthicalImplications.pdf)

2a. Assessment of the Extended Proposal – Supervisor

The supervisor and second marker should complete the appropriate section of this form. The weighting for each aspect (e.g. 35 for background research) is agreed between the student and supervisor, and can only be altered subsequently by providing a justification. The weights must stay within the ranges given and must add up to 100. Assign a mark for each aspect (e.g. 25 out of 35 for background research) and add these up to give the total mark. **Justify your marks** by writing comments on each aspect for consideration by the examination board and External Examiners. **Separately**, provide an overall evaluation below the table as feedback for the student. Send the completed form to the programme administrator (pgadmin@dcs.bbk.ac.uk).

Supervisor:	Date returned:
--------------------	-----------------------

Aspects	Comments	Weight	Mark	Revised mark
Background research		35 (30-40)		
Presentation of the problem – aims and objectives		15 (10-20)		
Plan for developing the solution		30		
Presentation of the proposal		20		
Total	(Add justification for revised marks here if applicable.)	100		

Comments to student:

--

2b. Assessment of the Extended Proposal – Second marker

Second marker:	Date returned:
-----------------------	-----------------------

Aspects	Comments	Weight	Mark	Revised mark
Background research		35 (30-40)		
Presentation of the problem – aims and objectives		15 (10-20)		
Plan for developing the solution		30		
Presentation of the proposal		20		
Total	(Add justification for revised marks here if applicable.)	100		

Comments to student:

--

3a. Assessment of the Report – Supervisor

Complete the appropriate section below (supervisor, second marker, third marker). The supervisor must confirm the date that a running version of the software developed by the student was demonstrated to you. The weighting for each aspect (e.g. 30 for implementation) is agreed between the student and supervisor, and can only be altered subsequently by providing a justification. The weights must stay within the ranges given and must add up to 100. Assign a mark for each aspect (e.g. 20 out of 30 for implementation) and add these up to give the total mark. **Justify your marks** by writing comments on each aspect for consideration by the examination board and external examiners. **Separately**, provide an overall evaluation below the table as feedback for the student. Send the completed form to the programme administrator (pgadmin@dcs.bbk.ac.uk).

Supervisor:	Date returned:
<i>I confirm that a running version of software developed by the student has been demonstrated to me on the date shown.</i>	Date software demonstrated:

Aspects	Comments	Weight	Mark	Revised mark
Specification and design		20 (10-30)		
Implementation, or execution of research		30 (20-40)		
Testing, results, analysis, critical evaluation		30		
Presentation and completeness of report, documentation		20		
Total	(Add justification for revised marks here if applicable.)	100		

Comments to student:

--

3b. Assessment of the Report – Second marker

Second marker:	Date returned:
-----------------------	-----------------------

Aspects	Comments	Weight	Mark	Revised mark
Specification and design		20 (10-30)		
Implementation, or execution of research		30 (20-40)		
Testing, results, analysis, critical evaluation		30		
Presentation and completeness of report, documentation		20		
Total	(Add justification for revised marks here if applicable.)	100		

Comments to student:

--

3c. Assessment of the Report – Third marker

Second marker:	Date returned:
-----------------------	-----------------------

Aspects	Comments	Weight	Mark	Revised mark
Specification and design		20 (10-30)		
Implementation, or execution of research		30 (20-40)		
Testing, results, analysis, critical evaluation		30		
Presentation and completeness of report, documentation		20		
Total	(Add justification for revised marks here if applicable.)	100		

Comments to student:

--

4. Assessment Criteria

To **pass (at least 50%)** a project the markers assess whether the project proposal and project report meet the following criteria. They also assess any other aspect of special relevance for the project.

Project Proposal:

- *Background research*: Potential approaches are reviewed and evaluated.
- *Presentation of the problem – aims and objectives*: The proposal specifies a suitable problem, and discusses its requirements.
- *Plan for developing the solution*: A suitable development/research method is chosen. The project is broken down into manageable chunks.
- *Presentation of the proposal*: Assessed as for the report – see below.

Project Report:

- *Specification and design*: Before starting the implementation, a specification and design of the system/software is laid out.
- *Implementation, or execution of research*: The key stages of the implementation/research are explained. The implementation/research is sound.
- *Testing, results, analysis and critical evaluation*: The report attempts to provide a clear and justified reflection upon the contribution and its limitations. It discusses how the software meets the specified requirements. A running version of the software is demonstrated to the supervisor (and an executable/source code on CD/DVD is turned in with the report).
- *Presentation of the proposal/report and documentation*: These are coherent in style and structure. They clearly communicate the student's contribution to the reader.

For a **distinction (at least 70%)**, a student would have to attempt a challenging project (this should be discussed and agreed with the potential supervisor) and gain a high grade under each of the above headings. To award a distinction the markers assess the report according to the following criteria:

Project Proposal:

- *Background research*: The student shows a clear understanding of the researched material. Potential approaches are reviewed and critically evaluated, highlighting strengths and weaknesses of each.
- *Presentation of problem – aims and objectives*: A challenging problem is specified and clearly outlined: this includes its context and the technical/user requirements.
- *Plan for developing the solution*: An appropriate development/research method is chosen and its suitability is well-justified. The project is broken down into subtasks that are logically coherent. In the case of unknowns (e.g. open research questions) "fallback" plans are laid out.
- *Presentation of the proposal*: Assessed as for the report – see below.

Project Report:

- *Specification and design*: The specification and design of the system/software shows a clear understanding of what needs to be done to meet the requirements, and is well-rounded, i.e. the components fit together in a coherent way.
- *Implementation, or execution of research*: The key stages of the implementation/research are clearly explained. The implementation/research is done to a high standard.
- *Testing, results, analysis and critical evaluation*: The solution demonstrates real insight into the problem/research question. There is clear and justified reflection upon the contribution and its limitations. The key results are accurately analysed and their relevance is explained. It is discussed how the software meets the specified requirements and is shown to be reliable. The author critically assesses the results and draws relevant conclusions from the study. A running version of the software is

demonstrated (as above).

- *Presentation of the proposal/report and documentation*: Complex issues are explained clearly and concisely. The content is well-organised and structured in a way that demonstrates the links between the concepts presented. The proposal/report shows that the student clearly understands the researched material. The solution and any other claims made by the students are well-justified. The author uses various resources and cites relevant resources using an appropriate consistent referencing style. The proposal/report is of professional quality and contains very few, ideally no, typographic errors.

Work that meets some, but not all, of the criteria for distinction may be considered for a **merit (between 60% and 69%)**. A merit might be awarded for a respectable, if only partially successful, attempt at a challenging project, or for a less ambitious project carried out, and written up, to a high standard.

The separate examiners grade the project independently and then meet to arrive at an agreed grade. Students may be called upon to make a presentation of their projects to a sub-committee of the Examination Board to demonstrate their grasp of the material.