

## 12 CM3060 Natural Language Programming

### 12.1 Project Idea 1: Identifying research methodologies that are used in research in the computing disciplines.

**What problem is this project solving, or what is the project idea?**

**What is the background and context to the question or project idea above?**

All research should be guided by a process that begins with the researcher's philosophical world view and then details why the research has value, how the research was done, and why the particular approach was used. In addition, this process should be made known to the research community. The "how" and "why" are the chosen research methodology. The idea behind this project is to take published research (journal articles, theses, etc.) in the computing fields, determine the computing discipline (Computer Science, Information Systems, Information Technology, etc.), determine the field within the discipline and then determine what research methodologies are used in these publications.

**Here are some recommended sources for you to begin your research.**

D Jurafsky and J Martin. Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models, Third Edition draft, University of Colorado at Boulder

B. J. Oates. Researching Information Systems and Computing. SAGE, London, 2006.

C. Pilkington and L. Pretorius. A conceptual model of the research methodology domain. In Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2015) - Volume 2: KEOD, pages 96–107, Lisbon, Portugal, November 2015.

The Natural Language Toolkit <https://www.nltk.org>

**What would the final product or final outcome look like?**

A system that classifies research publications according to computing discipline, field and the research methodologies used in that publication.

**What would a prototype look like?**

Some progress on the way to achieving the above.

**What kinds of techniques/processes/CS fundamentals are relevant to this project?**

Natural Language Processing – Information Retrieval in particular, but the whole spread of NLP is in the frame for this project.

Depending on the approach taken, some Machine Learning techniques.

Some of the NLP 'black box' approaches as appropriate.

**What would the output of these techniques/processes/CS fundamentals look like?** A research methodology classification pipeline for the computing disciplines.

**How will this project be evaluated and assessed by the student (i.e. during iteration of the project)? What criteria are important?**

Depending on the approach taken, the relevant and appropriate metrics should be used to obtain feedback as to whether the approach is suitable, and there should be some points at which possible other approaches are considered and justified.

**For this brief, what might a minimum pass (e.g. 3rd) student project look like?**

A system that is able to take a publication as input, and is able to determine the computing discipline.

**For this brief, what might a good (e.g. 2:2 – 2:1) student project look like?** As above, but able to also identify field within the discipline.

**For this brief, what might an outstanding (e.g. 1st) student project look like?**

A complete working classification pipeline that gives useful output.