



Board of Studies

Course Proposal Template

PROPOSED COURSE TITLE: Natural Language Processing 2

PROPOSER(S): Adam Lopez, Rico Sennrich, Sharon Goldwater

DATE: 2 March, 2018

SUMMARY

This template contains the following sections, which should be prepared roughly in the order in which they appear (to avoid spending too much time on preparation of proposals that are unlikely to be approved):

1. Case for Support

– To be supplied by the proposer and shown to the BoS Academic Secretary prior to preparation of an in-depth course description

1a. Overall contribution to teaching portfolio

1b. Target audience and expected demand

1c. Relation to existing curriculum

1d. Resources

2. Course descriptor

- This is the official course documentation that will be published if the course is approved, ITO and the BoS Academic Secretary can assist in its preparation

3. Course materials

- These should be prepared once the Board meeting at which the proposal will be discussed has been specified

3a. Sample exam question

3b. Sample coursework specification

3c. Sample tutorial/lab sheet question

3d. Any other relevant materials

4. Course management

- This information can be compiled in parallel to the elicitation of comments for section 5.

4a. Course information and publicity

4b. Feedback

4c. Management of teaching delivery

5. Comments

- To be collected by the proposer in good time before the actual BoS meeting and included as received

5a. Year Organiser Comments

5b. Degree Programme Co-Ordinators

5c. BoS Academic Secretary

[Guidance in square brackets below each item. Please also refer to the guidance for new course proposals at <http://www.inf.ed.ac.uk/student-services/committees/board-of-studies/course-proposal-guidelines>. Examples of previous course proposal submissions are available on the past meetings page <http://web.inf.ed.ac.uk/infweb/admin/committees/bos/meetings-directory>.]

SECTION 1 – CASE FOR SUPPORT

[This section should summarise why the new course is needed, how it fits with the existing course portfolio, the curricula of our Degree Programmes, and delivery of teaching for the different years it would affect.]

1a. Overall contribution to teaching portfolio

[Explain what motivates the course proposal, e.g. an emergent or maturing research area, a previous course having become outdated or inappropriate in other ways, novel research activity or newly acquired expertise in the School, offerings of our competitors.]

The field of NLP has been changing rapidly in the last few years with the advent of neural network-based approaches. As a consequence, many areas of NLP that formerly used distinct models and algorithms are now more unified under the neural framework. This means that some of our level 11 semester 2 NLP courses now have much more overlapping content than previously, and in particular both NLU and MT include 2-3 weeks of introductory material on neural nets for students who haven't seen it elsewhere (a significant minority). We also note that a third NLP course, Natural Language Generation, has not run in at least five years due to resource limitations.

We therefore propose to rationalize the level 11 NLP curriculum by replacing NLU, MT, and NLG (each 10 credits) with a single 20-credit course, Natural Language Processing 2. The course will focus mainly (but not necessarily exclusively) on neural network approaches to NLP, and will include several different application areas that were previously covered in the three different courses (e.g., translation, summarization, semantic parsing).

1b. Target audience and expected demand

[Describe the type of student the course would appeal to in terms of background, level of ability, and interests, and the expected class size for the course based on anticipated demand. A good justification would include some evidence, e.g. by referring to projects in an area, class sizes in similar courses, employer demand for the skills taught in the course, etc.]

The course is mainly aimed at the many MSc students who come to Edinburgh to specialize in NLP. It will also be open to ug4/5 students. This year, there are 107 students in MT, 77% of whom are also taking NLU; and 147 students in NLU, 57% of whom are also taking MT. However NLP enrollments have been growing by ~30% per year for the last four years, and it is likely that another 10-credit NLP course (TNLP, 31 students) will not run next year. We therefore anticipate between 110-150 students in the new course.

1c. Relation to existing curriculum

[This section should describe how the proposed course relates to existing courses, programmes, years of study, and specialisms. Every new course should make an important contribution to the delivery of our Degree Programmes, which are described at http://www.drps.ed.ac.uk/17-18/dpt/drps_inf.htm.

Please name the Programmes the course will contribute to, and justify its contribution in relation to courses already available within those programmes. For courses available to MSc students, describe which specialism(s) the course should be listed under (see <http://web.inf.ed.ac.uk/infweb/student-services/ito/students/taught-msc-2017/programme-guide/specialist-areas>), and what its significance for the specialism would be. Comment on the fit of the proposed course with the structure of academic years for which it should be offered. This is described in the Year Guides linked from <http://web.inf.ed.ac.uk/infweb/student-services/ito/students>.]

This course is intended as a follow-on to either ANLP (for MSc students) or FNLP (for undergraduates). Together, these courses are the backbone of the MSc specialist area in NLP, which is part of the MSc degrees in AI, Informatics, and Cognitive Science. It is also relevant as an application area for Data Science. Similarly, the course contributes to UG programmes in AI and Cognitive Science.

(See also section 1a.)

1d. Resources

[While course approvals do not anticipate the School's decision that a course will actually be taught in any given year, it is important to describe what resources would be required if it were run. Please describe how much lecturing, tutoring, exam preparation and marking effort will be required in steady state, and any additional resources that will be required to set the course up for the first time. Please make sure that you provide estimates relative to class size if there are natural limits to its scalability (e.g. due to equipment or space requirements). Describe the profile of the course team, including lecturer, tutors, markers, and their required background. Where possible, identify a set of specific lecturers who have confirmed that they would either like to teach this course apart from the proposer, or who could teach the course in principle. It is useful to include ideas and suggestions for potential teaching duty re-allocation (e.g. through course sharing, discontinuation of an existing course, voluntary teaching over and above normal teaching duties) to be taken into account when resourcing decisions are made.]

We anticipate that by combining MT, NLU, and NLG into a single course, we will achieve more efficiency of resourcing; in particular, the organizational effort for labs and tutorials need not be duplicated, and only one exam will be needed. The course will have three lectures per week, and 3 lab sessions over the semester.

SECTION 2 – COURSE DESCRIPTOR

[This is the official course descriptor that will be published by the University and serves as the authoritative source of information about the course for student via DRPS and PATH. Current course descriptions in the EUCLID Course Catalogue are available at www.euclid.ed.ac.uk under 'DPTs and Courses', searching for courses beginning 'INFR']

2a. Course Title [Name of the course.]:

Natural Language Processing 2

2b. SCQF Credit Points:

[The Scottish Credit and Qualifications Framework specifies where each training component provided by educational institutions fits into the national education system. Credit points per course are normally 10 or 20, and a student normally enrolls for 60 credits per semester. For those familiar with the ECTS system, one ECTS credit is equivalent to 2 SCQF credits. See also <http://www.scqf.org.uk/The%20Framework/Credit%20Points>.]

20

SCQF Credit Level:

[These levels correspond to different levels of skills and outcomes, see http://www.sqa.org.uk/files_ccc/SCQF-LevelDescriptors.pdf At University level, Year 1/2 courses are normally level 8, Year 3 can be level 9 or 10, Year 4 10 or 11, and Year 5/MSc have to be level 11. MSc programmes may permit a small number (up to 30 credits overall) of level 9 or 10 courses.]

11

Normal Year Taken: 1/2/3/4/5/MSc

[While a course may be available for more than one year, this should specify when it is normally taken by a student. "5" here indicates the fifth year of undergraduate Masters programmes such as MInf.]

4

Also available in years: 1/2/3/4/5/MSc

Different options are possible depending on the choice of SCQF Credit Level above: for level 9, you should specify if the course is for 3rd year undergraduates only, or also open to MSc students (default); for level 10, you should specify if the course is available to 3rd year and 4th year undergraduates (default), 4th year undergraduates only, and whether it should be open to MSc students; for level 11, a course can be available to 4th and 5th year undergraduates and MSc students (default), to 5th year undergraduates and MSc students, or to MSc students only]

5, MSc

Undergraduate or Postgraduate?

[If the course is only available to MSc students, then it must be classified as a Postgraduate course. All other courses, regardless of level, will be classified as Undergraduate]

Undergraduate

2c. Subject Area and Specialism Classification:

[Any combination of Computer Science, Artificial Intelligence, Software Engineering and/or Cognitive Science as appropriate. For courses available to MSc students, please also specify the relevant MSc specialist area (to be found in the online MSc Year Guide at <http://web.inf.ed.ac.uk/infweb/student-services/ito/students/taught-msc-2017/programme-guide/specialist-areas>), distinguishing between whether the course should be considered as “core” or “optional” for the respective specialist area.]

AI, CG

Appropriate/Important for the Following Degree Programmes:

[Please check against programmes from http://www.drps.ed.ac.uk/17-18/dpt/drps_inf.htm to determine any specific programmes for which the course would be relevant (in many cases, information about the Subject Area classification above will be sufficient and specific programmes do not have to be specified). Some courses may be specifically designed for non-Informatics students or with students with a specific profile as a potential audience, please describe this here if appropriate.]

All UG and MSc degrees with AI, cognitive science, data science, or Informatics.

Timetabling Information:

[Provide details on the semester the course should be offered in, specifying any timetabling constraints to be considered (e.g. overlap of popular combinations, other specialism courses, external courses etc).]

Semester 2. Needs to avoid clashes with ML courses and core courses on the PPLS Speech and Language Processing degree.

2d. Summary Course Description:

*[Provide a brief official description of the course, **around 100 words**. This should be worded in a student-friendly way, it is the part of the descriptor a student is most likely to read.]*

This course explores current research on processing natural language: interpreting, generating, and translating. The course will focus mainly on deep learning approaches to various NLP tasks and applications. It will provide an introduction to the main neural network architectures used in NLP and how they are used for tasks such as syntactic and semantic parsing, as well as end-user applications such as machine translation and text summarization. Building on linguistic and algorithmic knowledge taught in prerequisite courses, this course also aims to further develop students' understanding of the strengths and weaknesses of current approaches with respect to linguistic and computational considerations. Practical assignments will provide the opportunity to implement and analyse some of the approaches considered.

Course Description:

[Provide an academic description, an outline of the content covered by the course and a description of the learning experience students can expect to get. See guidance notes at: http://www.studentsystems.is.ed.ac.uk/staff/Support/User_Guides/CCAM/CCAM_Information_Captured.html]

This course explores current research on processing natural language: interpreting, generating, and translating. The course will focus mainly on deep learning approaches to various NLP tasks and applications. It will provide an introduction to the main neural network architectures used in NLP and how they are used for tasks such as syntactic and semantic parsing, as well as end-user applications such as machine translation and text summarization. Building on linguistic and algorithmic knowledge taught in prerequisite courses, this course also aims to further develop students' understanding of the strengths and weaknesses of current approaches with respect to linguistic and computational considerations. Practical assignments will provide the opportunity to implement and analyse some of the approaches considered.

Pre-Requisite Courses:

[Specify any courses that a student must have taken to be permitted to take this course. Pre-requisites listed in this section can only be waived by special permission from the School's Curriculum Approval Officer, so they should be treated as "must-have". By default, you may assume that any student who will register for the course has taken those courses compulsory for the degree for which the course is listed in previous years.]

Please include the FULL course name and course code].

Students MUST have passed: [Accelerated Natural Language Processing \(INFR11125\)](#) OR ([Informatics 2A - Processing Formal and Natural Languages \(INFR08008\)](#) AND [Foundations of Natural Language Processing \(INFR09028\)](#))

Co-Requisite Courses:

[Specify any courses that should be taken in parallel with the existing course. Note that this leads to a timetabling constraint that should be mentioned elsewhere in the proposal. Please include the FULL course name and course code].

Students MUST also take: Introductory Applied Machine Learning (INFR11069) OR Machine Learning and Pattern Recognition (INFR11130) OR Machine Learning Practical (INFR11132)

Prohibited Combinations:

[Specify any courses that should not be taken in combination with the proposed course. Please include the FULL course name and course code].

Other Requirements:

[Please list any further background students should have, including, for example, mathematical skills, programming ability, experimentation/lab experience, etc. It is important to consider that unless there are formal prerequisites for participation in a course, other Schools can register their students onto our courses, so it is important to be clear in this section. Also be aware that MSc students are unlikely to have the pre-requisite courses, so alternative knowledge should be recommended. If you want to only permit this by special permission, a statement like "Successful completion of Year X of an Informatics Single or Combined Honours Degree, or equivalent by permission of the School." can be included.]

Programming skills at least at the level of Computer Programming for Speech and Language Processing are also required.

This course is open to all Informatics students including those on joint degrees. For external students where this course is not listed in your DPT, please seek special permission from the course organiser.

Available to Visiting Students: Yes/No

[Provide a justification if the answer is No.]

yes

2e. Summary of Intended Learning Outcomes (MAXIMUM OF 5):

[List the learning outcomes of the course, emphasising what the impact of the course will be on an individual who successfully completes it, rather than the activity that will lead to this outcome. Further guidance is available from

<https://canvas.instructure.com/courses/801386/files/24062695>

On completion of this course, the student will be able to

1. Identify and discuss the main linguistic, machine learning, and ethical challenges involved in the development and use of natural language processing systems.
2. Understand and describe state-of-the-art models and algorithms used to address challenges in natural language processing systems.
3. Design, implement, and apply modifications to state-of-the-art natural language processing systems.
4. Understand the computational and engineering challenges that arise in the use of different models for natural language processing, and discuss the pros and cons of different models for a given task.
5. Understand, design and justify approaches to evaluation and error analysis in natural language processing systems.

Assessment Information

[Provide a description of all types of assessment that will be used in the course (e.g. written exam, oral presentation, essay, programming practical, etc) and how each of them will assess the intended learning outcomes listed above. Where coursework involves group work, it is important to remember that every student has to be assessed individually for their contribution to any jointly produced piece of work. Please include any minimum requirements for assessment components e.g. student must pass all individual pieces of assessment as well as course overall].

Written exam 70%, coursework 30%

Assessment Weightings:

Written Examination: 70%

Practical Examination: 0 %

Coursework: 30 %

Time spend on assignments:

[Weightings up to a 70/30 split between exam and coursework are considered standard, any higher coursework percentage requires a specific justification. The general expectation is that a 10-point course will have an 80/20 split and include the equivalent of one 20-hour coursework assignment (although this can be split into several smaller pieces of coursework. The Practical Examination category should be used for courses with programming exams. You should not expect that during term time a student will have more than 2-4 hours to spend on a single assignment for a course per week. Please note that it is possible, and in many cases desirable, to include formative assignments which are not formally assessed but submitted for feedback, often in combination with peer assessment.]

24 hours, across 2 or 3 assignments.

Academic description:

[A more technical summary of the course aims and contents. May include terminology and technical content that might be more relevant to colleagues and administrators than to students.]

The course aims to familiarize students with recent research across a range of topics within NLP, mainly within the framework of neural network models, and with a focus on applications such as machine translation, summarisation, and semantic parsing. As an MSc-level course that assumes previous experience with NLP, it will discuss a range of different issues, including linguistic/representational capacity, computational efficiency, optimization, etc. There is no textbook for the course; readings will come from recent research literature.

Syllabus:

*[Provide a more detailed description of the contents of the course, e.g. a list of bullet points roughly corresponding to the topics covered in each individual lecture/tutorial/coursework. The description should **not exceed 500 words** but should be detailed enough to allow a student to have a good idea of what material will be covered in the course. Please keep in mind that this needs to be flexible enough to allow for minor changes from year to year without requiring new course approval each time.]*

The syllabus will change to reflect recent research, but an indicative syllabus would include the following topics, including models, datasets, evaluation methods, and (where appropriate) ethical considerations:

Neural networks and word representations for NLP

- * Neural network basics
- * Feedforward and recurrent models
- * Word and sub-word representations

Sequence-to-sequence models and machine translation

- * Sequence-to-sequence models for MT
- * Attention models
- * Linguistic issues in MT

Parsing and semantic processing

- * Syntactic parsing
- * Semantic role labeling
- * Semantic parsing and question answering
- * Summarization

Relevant QAA Computing Curriculum Sections:

[Please see <http://www.qaa.ac.uk/en/Publications/Documents/SBS-Computing-consultation-15.pdf> to check which section the course fits into.]

Speech & natural language processing, Artificial intelligence, Human-computer interaction, Machine learning

Graduate Attributes, Personal and Professional skills:

[This field should be used to describe the contribution made to the development of a student's personal and professional attributes and skills as a result of studying this course – i.e. the generic and transferable skills beyond the subject of study itself. Reference in particular should be made to SCQF learning characteristics at the correct level http://www.sqa.org.uk/files_ccc/SCQF-LevelDescriptors.pdf.]

Students will develop their skills in reading research papers and identifying pros and cons of different approaches. They will also learn to analyze and discuss results from their own implementations.

Breakdown of Learning and Teaching Activities:

[Total number of lecture hours and tutorial hours, with hours for coursework assignments.]

[The breakdown of learning and teaching activities should only include contact hours with the students; everything else should be accounted for in the Directed Learning and Independent Learning hours.]

The total being 10 x course credits. Assume 10 weeks of lectures slots and 10 weeks of tutorials, though not all of these need to be filled with actual contact hours. As a guideline, if a 10-pt course has 20 lecture slots in principle, around 15 of these should be filled with examinable material; the rest should be used for guest lectures, revision sessions, introductions to assignments, etc. Additional categories of learning and teaching activities are available, a full list can be found at:

http://www.euclid.ed.ac.uk/Staff/Support/User_Guides/CCAM/Teaching_Learning.htm

Lecture Hours: 30 hours

Seminar/Tutorial Hours: 0 hours

Supervise practical/Workshop/Studio hours: 3 hours

Summative assessment hours: 2 hours

Feedback/Feedforward hours: 2 hours

Directed Learning and Independent Learning hours: 163 hours

Total hours: 200 hours

You may also find the guidance on 'Total Contact Teaching Hours' and 'Examination & Assessment Information' at:

http://www.studentsystems.ed.ac.uk/Staff/Support/User_Guides/CCAM/CCAM_Information_Captured.html

Keywords:

[A list of searchable keywords.]

Natural language processing

SECTION 3 - COURSE MATERIALS

3a. Sample exam question(s)

[Sample exam questions with model answers to the individual questions are required for new courses. A justification of the exam format should be provided where the suggested format non-standard. The online list of past exam papers gives an idea of what exam formats are most commonly used and which alternative formats have been http://www.inf.ed.ac.uk/teaching/exam_papers/.]

See recent exams for MT and NLU.

3b. Sample coursework specification

[Provide a description of a possible assignment with an estimate of effort against each sub-task and a description of marking criteria.]

See recent courseworks for MT and NLU.

3c. Sample tutorial/lab sheet questions

[Provide a list of tutorial questions and answers and/or samples of lab sheets.]

See recent lab sheets for MT.

3d. Any other relevant materials

[Include anything else that is relevant, possibly in the form of links. If you do not want to specify a set of concrete readings for the official course descriptor, please list examples here.]

See readings for NLU and MT courses on their course web pages.

SECTION 4 - COURSE MANAGEMENT

4a. Course information and publicity

[Describe what information will be provided at the start of the academic year in which format, how and where the course will be advertised, what materials will be made available online and when they will be finalised. Please note that University and School policies require that all course information is available at the start of the academic year including all teaching materials and lecture slides.]

It will be advertised as usual in MSc handbooks. The course web page will be created for the start of AY 2018-19.

4b. Feedback

[Provide details on feedback arrangements for the course. This includes when and how course feedback is solicited from the class and responded to, what feedback will be provided on assessment (coursework and exams) within what timeframe, and what opportunities students will be given to respond to feedback.]

The University is committed to a baseline of principles regarding feedback that we have to implement at every level, these are described at http://www.docs.sasg.ed.ac.uk/AcademicServices/Policies/Feedback_Standards_Guiding_Principles.pdf.

Further guidance is available from <http://www.enhancingfeedback.ed.ac.uk/staff.html>.]

Some of the lecture time will be devoted to discussing questions, including some exam-like questions, and providing feedback on student answers. Students will also get feedback on their work through labs.

4c. Management of teaching delivery

[Provide details on responsibilities of each course staff member, how the lecturer will recruit, train, and supervise other course staff, what forms of communication with the class will be used, how required equipment will be procured and maintained. Include information about what support will be required for this from other parties, e.g. colleagues or the Informatics Teaching Organisation.]

The course lecturer(s) will manage course staff following current practices of NLU and MT. Communication will be through the course mailing list, learn, and a course Q&A forum.

SECTION 5 - COMMENTS

[This section summarises comments received from relevant individuals prior to proposing the course. If you have not discussed this proposal with others please note this].

The idea of replacing MT/NLU/NLG with a single course was discussed at an ILCC meeting in Dec which all of the academic staff in NLP attended, and was approved by consensus.

5a. Year Organiser Comments

[Year Organisers are responsible for maintaining the official Year Guides for every year of study, which, among other things, provide guidance on available course choices and specialist areas. The Year Organisers of all years for which the course will be offered should be consulted on the appropriateness and relevance on the course. Issues to consider here include balance of course offerings across semesters, subject areas, and credit levels, timetabling implications, fit into the administrative structures used in delivering that year.]

5b. BoS Academic Secretary

[Any proposal has to be checked by the Secretary of the Board of Studies prior to discussion at the actual Board meeting. This is a placeholder for their comments, mainly on the formal quality of the content provided above.]