University of Salford MANCHESTER

School of Computing, Science & Engineering

School of Computing, Science & Engineering

Assessment Briefing to Students

The Student Handbook for your programme contains information about assessment, plagiarism, handing-in procedure, marking of late work, claiming mitigating circumstances and the week numbers.

Assessment Title :	**Semantic Web and Information Extraction Assignment**				
Module Title :	**Semantic Web and Information Extraction**				
Module CRN Code :	34357 and 34906		**Level:** 7	**Semester:**	2
Programme Code(s):	MSc Databases & Web-based Systems / MSc Advanced Computer Science		**Issue date[1]:**	16/03/2017	
Weighting :	**50**	% of the total module mark	**Submission date[2]:**	**28/04/2017 by 4pm via Blackboard**	
Assessor(s) :	Professor Apostolos Antonacopoulos		**Return date[3]:**	**19/05/2017**	

Part 1 – Semantic Web for the UK Bird Watching Society

This part represents 50% of the overall coursework marks available for the module.

Description:

The UK Bird Watching Society wants to develop a Semantic Web containing information about the wildlife birds seen around the UK. Birds may be seen in built-up areas (towns), countryside, coastal areas and nature reserves.

The Society has collected three basic sources of unstructured information about each kind of bird:

- Colour photographs of each bird.
- Sound recordings of the typical noise made by the bird.
- Text describing the bird's natural habitat and other useful facts.

You are required to develop a small Semantic Web that contains information about the birds, and then perform typical queries that would be made by birdwatchers.

The stages involved are outlines below:

a) Bird selection (**10% of the marks**)
You should select 3-4 species for **each** of the four categories of habitat mentioned above, and consider the type of information that can be derived from each of the data sources to provide a rich description of each bird.

For example, the information of value in the photographs could be the size of each bird, the colour/texture of its wings and body. The relevant sound parameters could be pitch, loudness, duration, regularity etc. The information in the text description could include: wingspan, number of eggs laid, typical food, natural habitat, nocturnal/daytime, migration

[1] Date on which brief was given to students
[2] Date by which assessment is to be submitted
[3] Date by which feedback will be made available to students. This must be within 15 working days of the submission date but takes staff annual leave and university closure into consideration

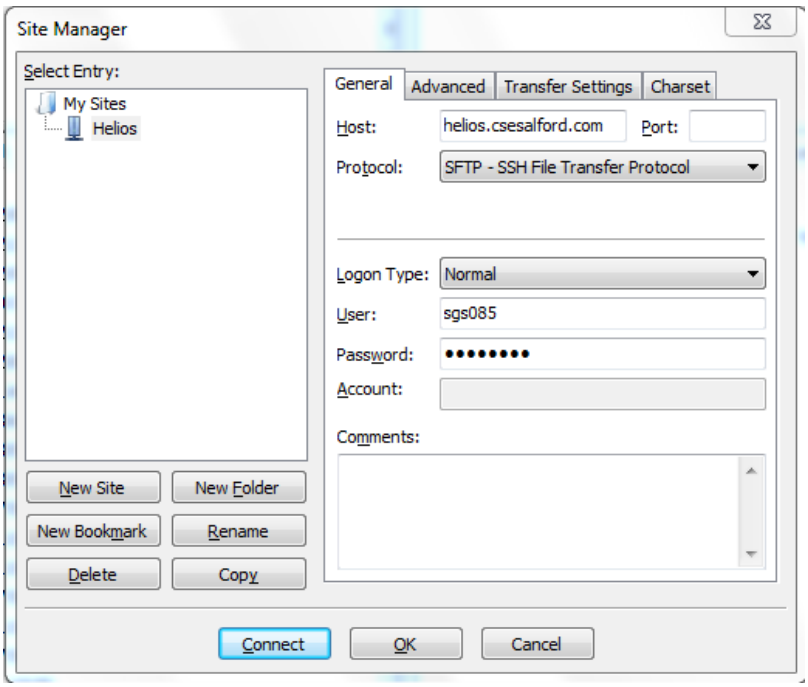
pattern etc. This information can all be taken from Wikipedia or the Royal Society for the Protection of Birds website, *rspb.org.uk.*

Generate (manually) a set of facts about each bird. In addition to simple facts you should also aim to have some *enumerated set values* e.g. size (small, median, large) and loudness (quiet, medium, loud) and *Boolean values* e.g. nocturnal/daytime; numeric values e.g. typical wingspan.

b) RDF description of the birds (**20% of the marks**)

Develop an RDF graph which contains assertional knowledge about individual animals as well as terminological knowledge (RDFS – classes and properties) for this specific application. Consider the use of URIs versus literals for the attributes from above and give a short explanation for your decisions. The RDF graph is to be serialised as RDF/XML. You can use the W3C validator (http://www.w3.org/RDF/Validator/) to verify the syntax and graph structure of your XML file.

Upload your RFD/XML file to your **Helios** web-space. If you cannot remember how to do this, you use **FileZilla** and select **File** then **Site Manager** and click the N**ew site** button. Complete the data entry screen as shown in the picture below, but using your *username* and **Helios** *password* and click the **Connect** button. You can then drag/drop your RDF file to your site.



c) Semantic Web (**20% of the marks**)

Use OpenLink Virtuoso (http://demo.openlinksw.com/sparql/ to run 5 queries on your Semantic Web and confirm the correct results are produced. The queries should cover a variety of features provided by SPARQL including at least one example involving terminological knowledge (i.e. individuals of a certain class).

Some examples are (depending on the actual data and to be translated into SPARQL syntax):

Where are swallows typically seen?
Which bird has a red breast?
Which birds are nocturnal and what do they feed on?
Which bird has the largest wingspan?
Which is the biggest costal bird?

You will need to put your web-space URL in the **Default Graph URI** text box, and cut/paste your SPARQL queries in the **Query Text** box. Make sure that you select *Retrieve Remote RDF data for all missing source graphs* from the **Sponging** options as shown in the screen dump below. You can run the query by clicking the **Run Query** button.

Further Instructions

The main deliverable for Part 1 will be a report that contains:

Section A

i) Your bird selection, and choice of attributes for each bird.

ii) Discussion about how you would derive your chosen attributes from the 3 data sources.

Section B

iii) Your RDF description of these entities and attributes (graph plus RDF/XML serialisation).

Section C

iv) 5 sample queries and the results returned with a short explanation.

Part 2: Research Study

This part represents 50% of the overall coursework marks available for the module.

The objective of this exercise is to undertake a piece of research on a topic of your choice in Information Extraction. Some possible topics are listed below, but you are not limited to these topics and you are free to choose other topics that interest you most:

- named entity recognition,
- sentiment analysis,
- semantic analysis,
- relation detection and classification,
- audio extraction,
- face detection/recognition,
- gesture recognition,
- document analysis and recognition
- machine learning for information classification and interpretation.

Your research should be written up as a report. You should familiarise yourself with the assessment criteria, which give a clear indication of the level and depth expected.

Essentially, you must provide evidence that you have chosen a reasonably demanding investigation, probably with a substantial technical foundation, and can demonstrate that you have researched this topic using a variety of means at your disposal (literature searches, journals, books, Internet searches, etc). You should be attempting to interpret the information gained from your research in a novel way, synthesising the ideas and concepts discovered. You should focus on the most recent/novel research carried out in the area.

Assessment Criteria

Objectives/motivations of study: An unambiguous title and clear statement of the objectives of the study. Objectives need to be realisable: not overly ambitious or too trivial in nature.

Introduction to area of study: Clear demonstration of application areas of the topic of study undertaken, setting the context of the area of study chosen within the field of Information Extraction.

Background reading/Literature review/References: Demonstration of wide reading encompassing relevant books, journal articles, conference proceedings, Internet resources, etc.

Quality of research undertaken/Depth of understanding demonstrated: Evidence of an understanding of relevant technologies, evidence of synthesis of material, novel/original approach to the study.

Quality of the layout of the report: clearly labelled sections, neatly formatted layout, good use of English.