

COMP7630 Web Intelligence and Its Applications
Semester 2, 2022-23

Group Term Paper / Group Development Project

1. What is the objective of the Group Term Paper / Group Development Project?

The group term paper's objective is to carry out a literature review on the techniques and applications of a particular area related to Web Intelligence.

The group development project's objective is to develop a software prototype which applies the web intelligence and related techniques learned in this course to some real-world data.

2. What should be covered and submitted for the Group Term Paper?

Each group is expected to prepare a literature review paper on one or more interconnected topics related to web intelligence. For example, it can be an overview of the latest development of some more recently developed computational methods related to Web Intelligence. Some examples are:

- word embedding methodologies (such as Word2vec, FastText, Glove) applied to Web Information Retrieval (e.g., for better modelling of documents, better consideration of web layout, ...), Recommender Systems (e.g., for friend/post/media recommendation in social media), Network Mining (e.g., for identifying community of users with similar interests/properties, etc.);
- large language models (such as GPT, Bert) with the same applications in the previous point;

- Evolutionary algorithms for Web Intelligence applications such as layout optimization;
- Estimation of Distribution Algorithms (a particular class of evolutionary algorithms);
- Methodologies for document clustering and/or classification;
- Topic modelling methodologies.

You can either choose one of the examples provided or propose any other reasonable subject.

To search for papers, you can try: <https://scholar.google.com/> or <https://arxiv.org/> or <https://arxiv-sanity-lite.com/>.

Some papers may have involved mathematics and concepts. Just skip them but, before skipping, see if the papers provide a good introduction or not. If yes, you can try to check with the papers they cited in the introduction section and see if you find them readable and interesting. The process can be iterated until you identify a set of suitable papers.

Note:

- *I suppose all group members should contribute to the identification of the topic of the term paper. Then, different group members can contribute to different sections of the term paper to make the content as rich as you can. Yet, there should be someone(s) to make sure that the term paper after integration should be consistent in terms of terminology and format.*
- *In order to get a high mark, each group is suggested to identify at least 3 x [group size] papers, read them in detail, describe the key ideas, compare and contrast them, and provide your own views regarding new opportunities of application. The selected papers that you read and mention should be included in the reference section of your group term paper and in-text citations should be included throughout your report.*

Deliverables:

i) Group Term Paper

(submit softcopy using the filename convention “WI_GroupX_TermPaper.pdf”)

The group term paper should cover what being mentioned above, and can be structured as:

- Term Paper Title
- Group members' IDs and names
- Abstract
- Introduction of XXX (e.g., word2vec and related methods)
- Recent development of XXX
- Applications of XXX (e.g., news recommendation)
- Conclusion
- References

You should include a table at the end of your report to detail how different group members contribute to the project. The paper is expected to be between 10 and 15 pages (1.5 line spacing; 12pt font size; 2 cm for all margins). Title, group information, table of contents, and references are all excluded from the pages count. The paper may contain figures and tables, but do not abuse of them.

ii) Individual Report

(submit softcopy using the filename convention "IDXXXXXXXX_Report.pdf" one per student)

It should be a one-page document (1.5 line spacing; 12pt font size; 2 cm for all margins) covering your proposal of some new application or extension of the methods covered in the term paper. Novelty regarding the application is what being emphasized here (that is something of your own and not mentioned by others before).

iii) Video Presentation

(submit a link to a MP4 video one per group)

A video presentation where you present the work by using slides and talking over the slides. The video must be between 15 and 20 minutes in MP4 standard format. All of the group members are expected to talk in the video. You can upload it in OneDrive, Google Drive or any other cloud platform and provide an accessible link.

3. What should be covered and submitted for Group Development Project?

For the group development project, a number of steps are expected to be accomplished, including

- 1) **Problem identification** (what motivates you to acquire particular data from online sources; what insights to gain by analyzing the data)
- 2) **Data acquisition** from online sources (e.g., via web scrapping or APIs). E.g.,
 - a. Linkedin (people profiles; Linkedin API)
 - b. Github (development project details; Github API)
 - c. Spotify (music; Spotify Web API)
 - d. News articles (<https://newsapi.org/>)
 - e. Tripadvisor (trip information; Tripadvisor API)
 - f. IMDB (movies; web crawling)
 - g. Web sites of your choice with rich textual information but without API (web crawling)
 - h. Out-of-copyright books collected from the Gutenberg project (<https://www.gutenberg.org/>)
 - i. Simply download datasets from Kaggle (<https://www.kaggle.com/>) or any other website (in this case focus more on the algorithmic point-of-view)
- 3) **Data pre-processing** for cleaning up the data
- 4) **Data analysis** for gaining the insights
- 5) **Visualization of analysis result** through charts and/or tables for clear presentation

Some examples are:

- Automatic suggestion of hashtags using association rules mining (for Instagram, Twitter, Moodle, etc.);
- Characters importance through the chapters of narrative books (by using Named Entity Recognition + Social Network Analysis tools) obtained for example from the Gutenberg projects website;
- Building text sentiment classifier for posts and comments in social media networks;
- Compare different evolutionary algorithms (from libraries such as Nevergrad) to optimize classification and/or clustering of collected data.

- Evolutionary algorithms for optimizing the design of a web advertisement banner by simulated users feedback (if you build a simulation model, you can replace data acquisition part with simulation design and description).

You can either choose one of the examples provided or propose any other reasonable subject.

Note:

- *I suppose all group members should contribute to the problem identification step and the data source identification. Then, different members can be responsible for different parts of the project.*
- *More novel/innovative applications you can come up with, the higher grade you will get.*

Deliverables:

i) Group Project Report

(submit softcopy using the filename convention “WI_GroupX_ProjectReport.pdf” one per group + a ZIP file containing code and data used for the project using the filename convention “WI_GroupX_software.zip”)

The report should cover the details of all the steps you have carried out as well as the discussion on the results obtained. It can be structured as:

- Group Project Title
- Group members’ IDs and names
- Abstract
- Motivation
- Details of all the steps
- Discussion of the results obtained and your interpretation
- Conclusion
- References

You should include a table at the end of the report to detail how different group members contributed to the project. There is a page limit of 10 pages (1.5 line spacing; 12pt font size; 2 cm for all margins). Title, group information, table of contents, and references are all excluded from the pages count. The paper may

contain figures and tables, but do not abuse of them. The page limit means that also less than 10 pages may be enough. In fact, it is important that you describe every operations you made in a complete but concise way.

iii) Individual Report

(submit softcopy using the filename convention "IDXXXXXXXX_Report.pdf" one per student)

It should be a one-page report (1.5 line spacing; 12pt font size; 2 cm for all margins) including your own *insights gained, lessons learned* as well as *new possibilities* that you can think of after carrying out this group project.

iv) Video Presentation

(submit a link to a MP4 video one per group)

A video presentation where you present the work by using slides and talking over the slides. The video must be between 15 and 20 minutes in MP4 standard format. All of the group members are expected to talk in the video. You can upload it in OneDrive, Google Drive or any other cloud platform and provide an accessible link.

4. Evaluation

Group Project Report / Term Paper	75%
Video Presentation	15%
Individual Report	10%