

AutoSurvey: Extractive Multi-Document Text Summarization for Generating Literature Surveys

Shane Rhoads
email

Kobee Raveendran
kobee.raveendran@knights.ucf.edu

1. PROBLEM STATEMENT

Literature surveys are a valuable resource for researchers exploring a foreign technical domain. They serve as a centralized collection of information on recent or foundational works in a digestible manner, and give the reader a solid conceptual, comparative understanding of the works discussed.

However, obtaining a comprehensive compilation of influential research requires extensive manual effort, and intuitively portraying the numerous concepts and methods requires vast domain-specific knowledge. For this reason, literature surveys can be scarce, particularly in new or unpopular research areas.

To alleviate the burden of manual survey creation, we seek to develop a system capable of automatically generating meaningful literature surveys. The problem will be formulated as an extractive multi-document text summarization task.

This is an introduction part, in which you should include a clear motivation on the problem being addressed in this project (why should I care?). You will need to also define the problem in broad terms so that you can outline the motivation.

As the motivation is made clear, you want to also state the problem more formally; in terms that an NLP expert will understand.

2. RELATED WORK

At this stage, you don't want to provide a comprehensive list of related work, but rather you want to consider this as a part in which you will provide a list of the resources that will be used to assist you conducting the project. Find out some of the related work that would be relevant to this project, and summarize how similar or different your work to them is. Even better, highlight in broad terms what would the δ you think you will achieve by this magnificent work be.

3. HOW TO TRAIN YOUR DRAGON

To generate more detailed multi-document summaries, we

considered expanding on existing extractive multi-document summarization methods. Extractive summarization entails selecting key sentences from the source texts, and using them directly in the summary. In our approach, we must first segment documents by topic. Then, sentences containing prevalent topics can be ranked and selected for the final summary. Due to the nature of summarization methods, we will only be considering conceptual content in the form of sentences. Other facets of papers, such as figures, tables, and formulas, may not have their relevance accurately judged automatically, and will thus be disregarded in preprocessing.

Here should be the actual proposal. What methods are you proposing? Describe the method you will use in designing your training method for your dragons. Relate to the problem statement. The description should be specific so that reproduction of the training method you used for your dragon are reproducible. Avoid copying others' work and others' style, and be creative.

4. EVALUATION

In this section, you want to describe two things: the data that you will use for evaluating the method against the problem stated above, and the results. As for data, describe your source of dragons, in as much details as needed, but not too much that I won't have the time to read. Be realistic.

Here, I know that you won't have results, so don't worry. Describe to me the evaluation metrics that you will use for evaluating the approach on the dataset above. Describe concisely the steps you propose to use for evaluation against those methods/metrics. The description should be intended for the non-expert so that she is able to reproduce the results of your work.

A bonus would be if you could propose to compare your work against a baseline.

5. EXPECTED OUTCOMES AND RISKS

Here is your chance to tell me what you expect of outcomes.

Also you want to tell me what are the risks associated with the project, and how you plan to deal with them.

6. PLAN AND ROLES OF COLLABORATORS

Divide your project into components, and tell me who is going to work on what, how much time each will work on each item. Have a timeline for the project. Tasks may include coding, testing, evaluation and analysis, write-up, presentation, etc.

7. REFERENCES

This space would be your chance to list the resources you used for the proposal.

8. REFERENCES