Outlines:

Introduce the system in this project.
What’s the motivations?
What’s the challenges?
What’s the project about: STER.STM/LDA

To **perfect** and **dig deeper** about the system that is able to evaluate how accurate a title of a document describes its content, furthermore, based on the previous evaluation the system is enabled to recommend a better title which describes its content better than the original title.
• A title evaluation system based on topic models.
• LDA is a latent variable model of documents, where a document is regarded as a mixture of K latent topics, each of which is probability distribution over words.
• STM? A upgraded version of LDA, take document structure into account (sentences), ideally more accurate than LDA.

Notes: STM is more accurate than LDA in the sense of topic modeling, but the task here is title evaluation.
Current structure of the system:

- Data preprocessing, sentence splitter (Perl)
- Data preprocessing (Java)
- Topic model STM/LDA (C)
- Semantic similarity calculation (C)
- GEV distribution fitting (R)
- Evaluation results
Motivations:

• Ask yourself that how many times you were attracted by a title, but after you read through the whole article, you found that’s not what you were expected at all.

• Think about what happens if you will never be cheated by title.
Motivations:

• Ask yourself that how many times you were attracted by a title, but after you read through the whole article, you found that’s not what you expected for at all.

• Think about what happens if you will never be cheated by title.

• Answer? Maybe the today’s accuracy of online search will be significantly improved.
Challenges:

• This project requires a lot of knowledge from different fields.
• The current version of the system was built by several languages – Perl, Java, C and R.
• Enormous data to handle when dealing with a series of experiments.
Question time!

My email: u5199740@anu.edu.au