Introduction

- **Survey Generation**
  - A simple paragraph for introduction.
  - A Wikipedia-style survey includes multiple sections.
- **Problem Definition**
  - Abstractive summarization from a list of related input documents.
  - Generate short summary for EACH individual section.
- **Related work for Wikipedia page generation**
  - Generating the initial leading paragraph of a Wikipedia page (Liu et al., 2018; Liu and Lapata, 2019; Perez-Beltrachini et al., 2019).
- **Challenges**
  - No existing data: surfer100 (100 manually written SURveys From wEb Resources on scientific topics) for testing purposes.
  - Selecting and cleaning web page: heuristics with manual checking.
  - Long input sequence: two-stage method.
- **Contributions**
  - A two-stage method for generating Wikipedia-like surveys for scientific topics;
  - Surfer100 dataset for survey generation using web resources.

Methodology

- **Step 1: Content selection**
  - Not every single sentence is considered to be relevant.
  - Long input issue: rank all sentences with Semantic Search, WikiCite and RoBERTa-Rank.
- **Step 2: Abstractive Summarization**
  - Pre-trained models for generating abstractive summarization for each section; Hiersumm and BART.

Evaluation

- Human Evaluation:
  - Randomly select 20 concepts and ask two human judges to give scores (range 1-5) on the following four perspectives: readability, relevancy, redundancy and hallucination.

Conclusion

- **A two-stage method for generating Wikipedia-like surveys for scientific topics**;
- Surfer100 dataset for survey generation using web resources;
- Future work will be on collecting a larger corpus for survey generation, as well as a more efficient annotation pipeline;
- Advanced models for improving summarization quality.

**Surfer100 Corpus**

- Generating Wikipedia-style surveys from web resources:
  - History
  - Main Idea
  - Variations
  - Applications

**Fig 1. Comprehensive survey sections.**

**Tab 2. Sample model generated survey on the topic “text summarization”**

**Fig 2. Sample URLs in Surfer100.**

**Fig 3. Sample Survey in Surfer100.**

**Fig 4. Sample Scientific Concepts in Surfer100.**