

Rapport sur l'outil *summvis*

SUMMVIS: Interactive Visual Analysis of Models, Data, and Evaluation for Text Summarization

Sommaire :

Introduction

A - Architecture système

B - Fonctionnalités

C - Visualisation

Summvis est un outil qui propose une interface de visualisation interactive qui fournit une comparaison des systèmes de résumé, permettant une analyse fine des modèles, des données et des différentes métriques d'évaluation. Il propose des alignements sémantiques et lexicaux ainsi qu'une taxonomie des relations textuelles.

Installation : <https://github.com/robustness-gym/summvis?tab=readme-ov-file>

Nécessite de downgrade python à la version 3.9 et pandas à la version 1.4.1.

Article : <https://arxiv.org/pdf/2104.07605>

Dans un premier temps, nous présentons l'architecture système utilisé (A), dans un second temps les différentes fonctionnalités (B) et enfin une description de l'interface (C).

A - Architecture Système

SUMMVIS est implémenté comme une application Streamlit avec des composants HTML/JavaScript avec des fonctionnalités personnalisés pour enrichir l'interaction utilisateur. L'outil est préchargé avec des prédictions de modèles d'état de l'art sur des ensembles de données de référence et peut être adapté pour analyser des données personnalisées via l'API des jeux de données HuggingFace.

L'architecture de Streamlit permet d'écrire des applications de la même manière qu'un script Python simple, ce qui permet une facile prise en main sans compétences en front nécessaires.

B - Fonctionnalités :

- **A . Modele analysis** : On compare le document source et le résumé généré. Avec cette comparaison, cette fonctionnalité nous permet de comprendre la capacité du modèle à abstraire et à conserver fidèlement les informations présentes dans le document. Les différents outils

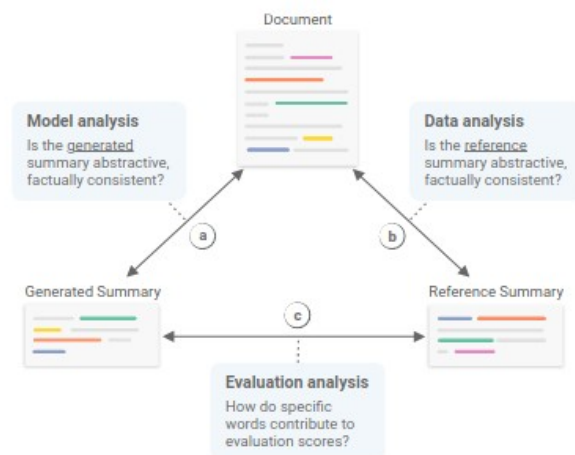


Figure 1: SUMMVIS supports fine-grained comparison between (a) source document and generated summary, (b) source document and reference summary, and (c) generated summary and reference summary, enabling analysis of models, data, and evaluation metrics.

permettent de mettre en évidence les différents éléments que les résumés partagent que ça soit sémantiquement ou lexicalement.

- **B . Data analysis** : On compare le document source avec le résumé de référence, cette fonctionnalité aide à déterminer dans quelle mesure le résumé de référence est lui-même abstrait et factuellement cohérent avec le document source. Tout comme l'analyse de modèle, ce sont les différents outils de l'application qui permettent une visualisation des relations entre les documents.
- **C . Evaluation analysis** : En comparant le résumé de référence avec le résumé généré, cette fonctionnalité met en lumière les relations entre les mots et les phrases qui forment la base des métriques d'évaluation automatique comme ROUGE et BERTScore. Ce niveau d'évaluation est différent des 2 premiers et permet d'avantage de comprendre les scores et de regarder précisément comment et pourquoi ils peuvent varier.

Voici plus précisément les fonctionnalités que propose l'application :

- **Alignement Lexical** : SUMMVIS identifie et souligne les n-grammes partagés entre les textes, facilitant ainsi la détection des passages extraits directement du document source.
- **Alignement Sémantique** : En utilisant les embeddings de mots, SUMMVIS évalue la similarité sémantique entre les textes, même lorsque les mots spécifiques ne correspondent pas exactement.
- **Taxonomie des Relations Textuelles** : SUMMVIS classe les relations textuelles selon une taxonomie qui inclut l'extraction, l'abstraction, l'hallucination et l'interprétation erronée, aidant ainsi à comprendre les comportements des modèles. C'est en considérant les mesures de similarités lexicales et sémantique que l'on peut proposer un classement.

1. *Extraction* : Haute similarité lexicale et sémantique. Le résumé cite textuellement le texte du document.
2. *Abstraction* : Faible similarité lexicale mais haute similarité sémantique. Le résumé consolide et paraphrase les informations du document.
3. *Hallucination* : faible similarité lexicale et sémantique. Le résumé est factuellement incohérent et comprend des informations qui sont absentes du document.
4. *Misinterpretation* : similarité lexicale élevée et faible similarité sémantique. Le résumé interprète mal et utilise des informations du document, telles que des homonymes mal compris.

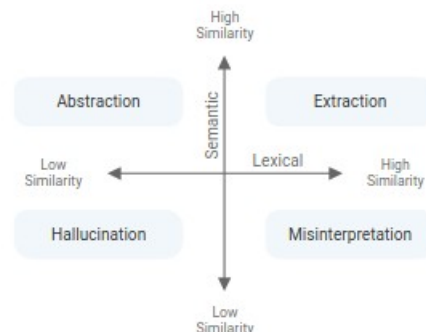


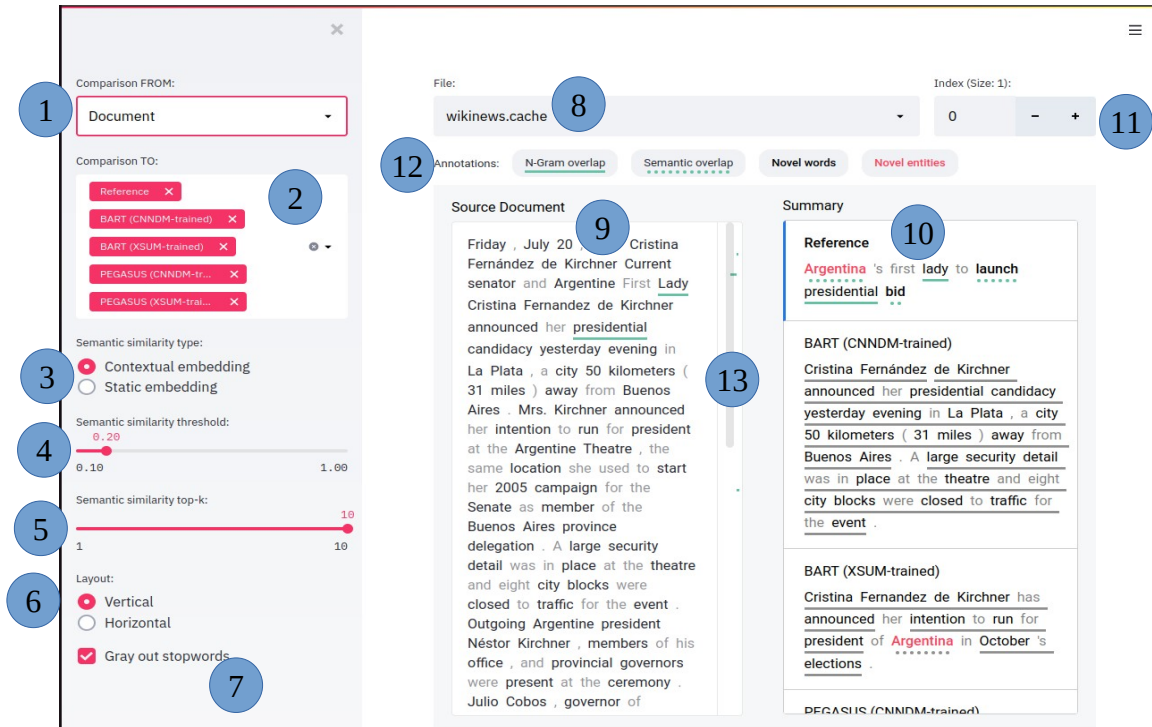
Figure 3: Taxonomy of textual relationships across lexical and semantic dimensions.

Les différentes annotations : (disponible dans la barre annotations, n°12 ci-dessous)

- **N-gram overlap**: séquences de mots que l'on retrouve dans le document source (ou le résumé de référence) et le résumé sélectionné à droite (se référer à la visualisation pour interpréter les indications). Les différentes couleurs utilisées pour soulignés sont définies selon l'index de la phrase du résumé.
- **Semantic overlap**: Mets en évidence dans le résumé les mots sémantiquement proches d'un ou plusieurs mots dans le document à gauche (c'est à dire le document source ou le résumé de référence). Les embeddings utilisés pour cet outil sont soit des embeddings contextuels (entraînés sur RoBERTa) soit des embedding static (spaCy).
- **Novel words**: Affiche les mots dans le résumé qui n'apparaissent pas dans le document avec lequel on le compare.

- **Novel entities:** Entités nouvelles qui n'apparaissent pas dans le document avec lequel on le compare (celui de gauche donc)

C - Visualisation :



Dans la barre de recherche de gauche :

- 1 - À partir de quoi on veut faire la comparaison (document ou résumé de référence) avec un menu déroulant.
- 2 - Quels modèles de résumé automatique on test, plusieurs modèles sont disponibles, on peut en mettre qu'un seul ou plusieurs.
- 3 - type de similarité sémantique : contextuelle ou statique
- 4 - curseur pour ajuster le seuil de similarité sémantique (pour le mot ou pour la phrase avec Semantic similarity threshold)
- 5 - Un curseur permettant de définir le nombre maximal de résultats de similarité sémantique à afficher (Semantic similarity top-k)
- 6 - Layout : Vertical/Horizontal, pour définir un type d'affichage
- 7 - Gray out stopwords : permet de mettre en gris les stopwords qui ne sont pas pris en compte par les différentes métriques.

Main :

- 8 et 11 - file/ Index
- 12 - Annotations : Des onglets pour afficher des informations sur le chevauchement de n-grammes, les nouveaux mots, et les nouvelles entités dans les résumés par rapport au document source. (Possible d'activer et de désactiver)
- 9 - Source document : document source
- 10 - Summary : Les différents résumés générés s'affichent dans une petite fenêtre.
- 13 - Barre de recherche avec une vue globale des annotations.