

# Seminar Sprachverarbeitung: Review

## 1. Titel der gereviewten Arbeit:

Exploring automatic extractive text summarization using the open text summarizer as an example

## 2. Autor der gereviewten Arbeit:

Anonym

## 3. Worum geht es in der Arbeit? Einordnung als Rezension/Survey/Vertiefung?

*Was ist das Thema und was ist die Leitfrage? Nenne die wesentlichen Punkte der Bearbeitung.*

Die Arbeit behandelt Verfahren zur automatischen Erzeugung von Zusammenfassungen von Texten. Eine konkrete Leitfrage besteht nicht; jedoch dient die Darstellung des grundlegenden Ansatzes (auszugsbasierter Zusammenfassungen) als roter Faden.

---

The paper examines techniques for generating summaries of texts automatically. There is no definite central question, but the illustration of the overall approach (extraction based summaries) serves as central theme.

## 4. Strukturierung der Arbeit, ist der rote Faden klar erkennbar? Wie ließe sich die Struktur verbessern?

*Einordnung der Arbeit in den Seminarkontext, Bezugnahme und Verweis auf andere behandelte Themen? (Tipp: Thema jedes Abschnitts an den Rand schreiben, dann prüfen ob die Struktur sinnvoll ist und zu den Abschnittsüberschriften passt.)*

Die Arbeit hat einen klar erkennbaren roten Faden. Beginnend mit der Vorstellung des Problems wird der behandelte Ansatz schrittweise teilweise mithilfe eines praktischen Beispiels vorgestellt, wobei zu jeder Zeit deutlich ist, weshalb der aktuelle Schritt sinnvoll ist und wozu er dient. Eine Einführung

liefert einen grundsätzlichen Zugang zur Problemstellung und das Resümee fasst die Arbeit noch einmal knapp zusammen und ordnet den Stand der verfügbaren Ansätze in die allgemeine Zielsetzung/Problemstellung ein. Eine grundsätzliche Verbesserung der Struktur halte ich nicht für notwendig.

---

The paper's central theme is very apparent. Partially being supported by a practical example, the approach being examined is presented step by step. At any time it is made understandable why the current step makes sense and which purpose it serves. In the beginning the basic problem statement is introduced and at the end the whole paper is being roughly summarized by the conclusion, which also gives some insight how the results achieved with the presented approach are comparing in matters of the overall goal. I do not consider a fundamental improvement to the papers structure to be necessary.

## 5. Welche Fragen bleiben nach dem Lesen der Arbeit offen? Passt das

*Resümee im Schlussteil der Arbeit?*

Beim Lesen der Arbeit stellt sich die Frage, wie sich der vorgestellte Ansatz (Extractive Summarization) zum Erstellen von Zusammenfassungen gegenüber alternativen Ansätzen einordnet - gibt es überhaupt andere ernstzunehmende Ansätze, wenn ja, wieso fiel die Wahl auf diesen? Das Resümee schließt die Arbeit sehr gut ab und erleichtert durch die darin enthaltene knappe Zusammenfassung das Verständnis des Gelesenen.

---

While reading the paper the following question occurs: How does the presented approach (extractive summarization) compete with regards to alternative approaches - are there any relevant alternatives at all? If there are, why was the extractive summarization approach chosen as the papers topic? The conclusion does a good job of completing the paper and helps to easier understand the article, by roughly summarizing it.

## 6. Einschätzung der Korrektheit vorgestellter Fakten und der

*Argumentationsketten: Finde Fehler (dann findet sie der Seminarleiter nicht mehr)!\**

Die vorgestellten Fakten und die Argumentation der Arbeit scheinen im Allgemeinen korrekt und schlüssig zu sein.

---

The presented information as well as the papers reasoning appears to be correct and coherent in general.

## 7. Anmerkungen zum Sprach- und Schreibstil, Grammatik,

*Rechtschreibung, etc.:*

Sprachlich gelingt dem Author die Balance zwischen inhaltlicher Dichte und Verständlichkeit sehr gut. Der Text richtet sich in erster Linie an Leser, die mit dem Forschungsgebiet noch nicht allzu vertraut sind. Dennoch vermittelt er ein sehr genaues Verständnis von der Herangehensweise, ohne zu viel Wissen vorauszusetzen. Die einzige Kritik an der Sprache der Arbeit stellen die gelegentlichen Rechtschreibfehler da.

---

The author does a good job of keeping the right proportion of information density to comprehensibleness. Despite being primarily meant for readers, who aren't too familiar with the field of study, the text imparts a reasonably detailed comprehension of the approach without requiring too much knowledge in the first place.

## 8. Formalia: Literaturverzeichnis, Zitierung, Formatierung, etc.

Die Zitate scheinen vollständig und formal richtig angegeben zu sein.

---

The quotes seem to be complete as well as formally correct.

## 9. Was sind die stärksten Aspekte/die größten Stärken des Papers?

Die größte Stärke der Arbeit liegt in der klaren Struktur, die auch praktische Beispiele miteinbezieht. Dies trägt dazu bei, dass das Paper gut zu lesen und leicht zu verstehen ist.

---

The papers biggest strength is it's clear structure which includes practical examples. This adds to the papers readability and comprehensibility.

## 10. Was sind die schwächsten Aspekte/die größten Schwächen des Papers?

*Wie lassen sich diese beheben?*

Die größte Schwäche des Papers ist die Abwesenheit einer 'echten' Leitfrage. Am ehesten erfüllt diese Funktion die Evaluation der Zusammenfassung, welche mithilfe des Open Text Summarizer für die Arbeit erstellt wurde und die Optimierung des Ergebnisses mithilfe der vorgestellten Methoden und Ansätze.

Diese ‘Leitfrage’ unterstützt zwar den roten Faden des Papers, sie hat jedoch außerhalb dessen keine Relevanz und ist deshalb auch als Motivation, die Arbeit zu lesen, ungeeignet. Es ist jedoch schwierig, das eigentliche Ziel - die Darstellung des Ansatzes zur automatischen Erzeugung von Zusammenfassungen - als Leitfrage zu formulieren, deshalb kann ich an dieser Stelle keinen guten Verbesserungsvorschlag machen.

---

The papers biggest weakness is the absence of a ‘real’ central question. To some extent the evaluation of the summarization, which was created as practical example using the Open Text Summarizer might fill that role. It supports the central theme and structure of the paper, yet it has no further relevance for the topic which makes it unsuitable to serve as motivation to read the paper. But to the authors defense: It is difficult to express the actual goal of the paper as central question, which is the presentation of the approach of generating extraction based summaries automatically. For this reason I struggle to propose an improvement here.

## 11. Sonstige Anmerkungen:

Im Fazit bewertet der Author die automatisch erzeugte Zusammenfassung durch die (satzweise) Übereinstimmung mit dem tatsächlich geschriebenen Abstract. Ich halte diese Bewertungsmethode nicht für sehr hilfreich, da das (tatsächliche) Abstract auf einer völlig anderen Grundlage (semantisches Verständnis der Aussage des Textes) beruht und im Gegensatz zu dem vorgestellten Verfahren zum Auszugsbasierten Erstellen von Zusammenfassungen Sätze enthalten kann, die nicht im Artikel vorkommen, bzw. sogar mehrere Sätze in einem zusammenfassen könnte. Sinnvoller, als die beiden Zusammenfassungen satzweise (nach Gleichheit) zu vergleichen wäre es, die Aussagen zu extrahieren und hier nach Übereinstimmung zu suchen. Dabei würde die Bewertung nicht mehr  $\text{recall}=0$ ,  $\text{precision}=0$  ergeben, sondern ein Ergebnis, dass den tatsächlichen Wert der Zusammenfassung besser widerspiegelt.

---

The author evaluates the automatically generated summary by comparing it (sentence by sentence, checking for equality) to the actually submitted abstract. I don’t consider this criteria to be very helpful, because these two summaries are created totally differently (semantic understanding of the text compared to the extraction of relevant sentences). For that reason the final abstract can contain sentences which are not taken from the article and maybe even conclude multiple sentences at once. Therefore I propose to extract the basic statements contained in both abstracts and compare them for semantic similarity. Als this would’n result in a evaluation of  $\text{recall}=0$  and  $\text{precision}=0$ , but a value that would represent the actual value of the summarization much better.

## **12. Anmerkungen an den Seminarleiter (werden nicht dem Autor der gereviewten Arbeit mitgeteilt):**

Das Problem mit der schwer zu formulierenden Leitfrage (siehe 10.) liegt, denke ich, in der Natur der Anforderung an die Seminararbeiten (Darstellung eines Gebiets aus der Sprachverarbeitung). Deshalb bin ich mir nicht sicher, wie relevant der Punkt in diesem Fall ist.

## **13. Anhang: konkrete Korrektur-/Verbesserungsvorschläge:**

*entweder der Form: "Abschnitt 2, zweiter Absatz, dritte Zeile:  
hier könntest Du wieder auf Graphik 2 verweisen." oder annotiertes PDF bzw.  
Scan anhängen.*

**siehe PDF**