

Von BLEU bis METEOR: Drei Jahre der Fortschritte in der Maschinellen Qualitätssicherung

Diese Proseminararbeit beschäftigt sich mit dem Vergleich der drei maschinelle Qualitätssicherungsverfahren BLEU sowie NIST und METEOR, welche auf BLEU basieren und dieses erweitern. Es wird darauf eingegangen, in welcher Weise BLEU die Qualitätssicherung ermöglicht und inwiefern NIST und METEOR dieses Verfahren weiter verbesserten. Zuerst werden die relevanten Begriffe der maschinellen Qualitätssicherung definiert um die Funktionsweise der drei Verfahren zu verstehen sowie einen allgemeinen Überblick zu verschaffen. Im zweiten Kapitel wird auf den grundlegenden Algorithmus von BLEU eingegangen und welche Vorteile dieser über vorhergehende Verfahren hatte, aber auch auf dessen Limitation. Im dritten Kapitel wird BLEU schließlich jeweils um Funktionen durch NIST und METEOR erweitert und es wird auf diese Verbesserungen eingegangen. Am Schluss werden die Unterschiede sowie dessen Bedeutung erneut zusammengefasst. Ziel dieser Proseminararbeit ist es, die rasanten Neuerungen, welche durch die Einführung von BLEU ermöglicht wurden, aufzuzeigen.

Inhaltsverzeichnis

Einleitung

1 Theoretische Grundlage (Powers 2011, Turian 2003)

Grundbegriffe der maschinellen Qualitätssicherung

2 Funktionsweise von BLEU (Papineni 2002)

2.1 Vorteile von BLEU

Was macht BLEU besser als andere Verfahren?

2.2 Nachteile von BLEU

Was kann BLEU nicht identifizieren?

3 Erweiterung des Verfahren durch NIST und METEOR

3.1 NIST (Doddington 2002)

Das Mizeinbeziehen von N-Grammhäufigkeit

3.2 METEOR (Bannerjee 2005)

Die Beachtung der N-Grammposition

4 Zusammenfassung

Bibliographie

Literaturverzeichnis

Banerjee, Satanjeev/Lavie, Alon (2005) „METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments“, in: *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization* 1:1, 65-72.

Doddington, George (2002) „Automatic Evaluation of Machine Translation Quality Using N-gram Co-Occurrence Statistics“, in: *HLT '02: Proceedings of the second*

international conference on Human Language Technology Research 2:1, 138-145.

Papineni, Kishore/Roukos, Salim/Ward, Todd/Zhu, Wei-Jing (2002) „BLEU: a Method for Automatic Evaluation of Machine Translation“, in *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 40:1, 311-318.

Powers, M. W. David (2011) „Evaluation: From Precision, Recall and F-Factorto ROC, Informedness, Markedness & Correlation“, in: *Journal of Machine Learning Technologies* 2:1, 37-63.

Turian, Joseph P./Shen, Luke/Melamed, I. Dan (2003) „Evaluation of machine translation and its evaluation“, in: *Proceedings of Machine Translation SUMmit IX: Papers* 9:1, 1-8.