

KI-basierte Qualitätsbewertung von Produktionsnachrichten und wissenschaftlichen Veröffentlichungen zur Additiven Fertigung

Zielstellung

Die Text Daten Qualität (DQ) ist das Kernstück des Text Mining. Trotz vieler theoretischer Rahmenwerke fehlt eine automatisch messbare Methodik für die Qualität Bewertung, insbesondere für Textdaten. Textdaten haben eine komplexe Syntax, Semantik und Datenform, und auch die Bewertungskriterien sind vielfältig und subjektiv. Die Herausforderung besteht darin, das Konzept von Textdaten Qualität zu interpretieren, zu messen und per Computer zu bewerten.

Vorgehen

Die Vorgehensweise lässt sich unterteilen in,

- Definition und Auswahl der DQ Dimensionen
- Bewertung der DQ-Metriken
- Normalisierung der Metriken
- Umsetzung nach Programm
- Studie zu Anwendungsfällen
- Untersuchung des Beeinflussens von PDF Format
- Clustering durch K-Means

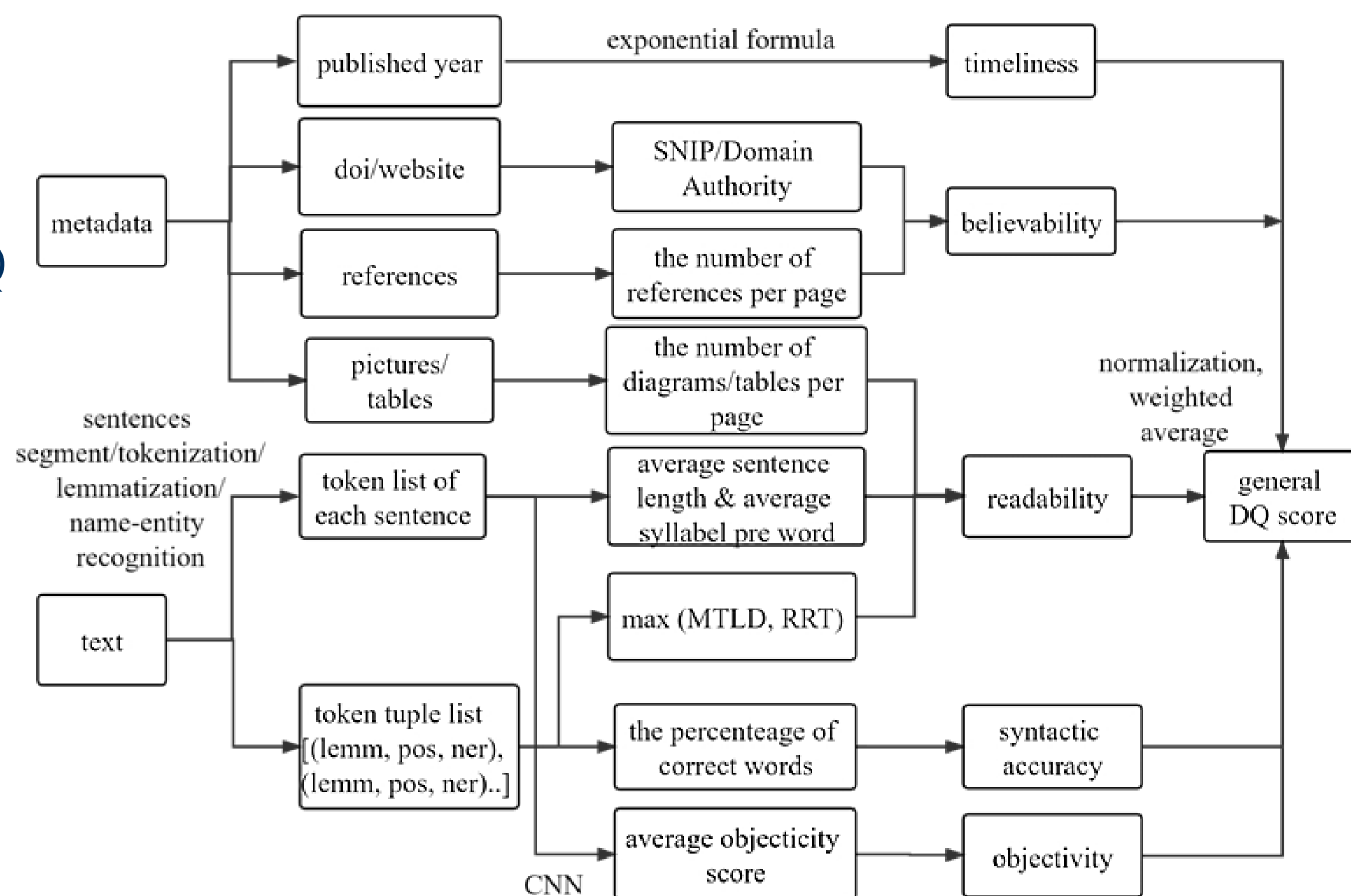


Abb.1 Der Umsetzungsansatz des Programms

Ergebnisse

In dieser Arbeit wurde eine sechs Dimension Methode zur Bewertung der Textqualität vorgeschlagen. Mehrere Metriken wurden ausgewählt, um die Bewertung zu quantifizieren und zu automatisieren.

Die Rechtschreibprüfung Hunspell für die syntaktische Genauigkeit und das CNN-Modell für die Objektivität zeigen gute Leistungen. Außerdem wurde nachgewiesen, dass das Datenformat einen großen Einfluss auf das Bewertungsergebnis hat. Für den PDF-Extraktor in der Publikation korrelieren die Ergebnisse der Textbewertung durch ocr2text und die manuelle Bewertung stark miteinander. Die diskrete Datenverteilung im Clustering zeigt, dass die Dimensionen insgesamt unabhängig sind.

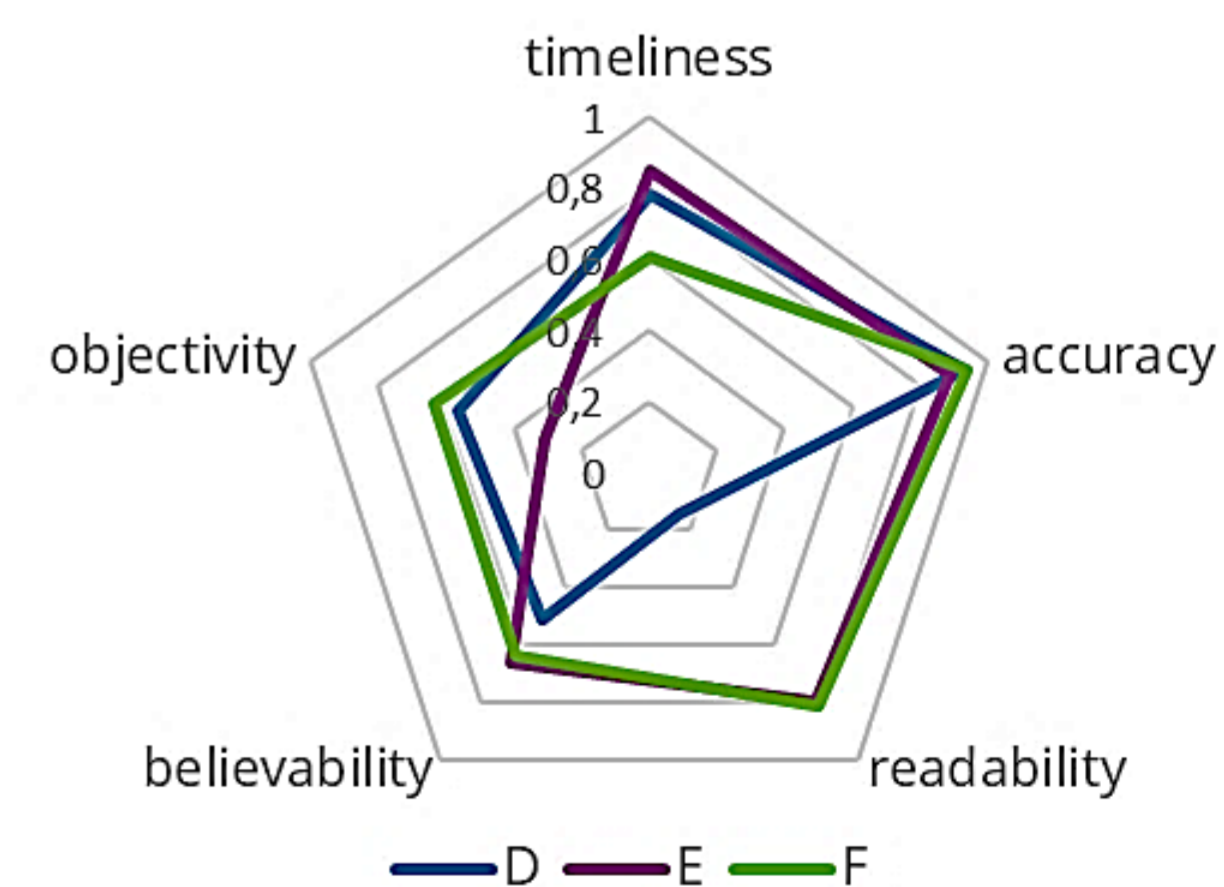


Abb.2 Beispiel-Radardiagramm von drei Texten