

# Entwicklung und Evaluierung eines Question Answering Systems im Rahmen des Cross Language Evaluation Forum (CLEF)

*Robert Strötgen, Thomas Mandl, René Schneider*

Universität Hildesheim  
Informationswissenschaft  
Marienburger Platz 22  
31141 Hildesheim  
mandl@uni-hildesheim.de

## **Zusammenfassung**

Question Answering Systeme versuchen, zu konkreten Fragen eine korrekte Antwort zu liefern. Dazu durchsuchen sie einen Dokumentenbestand und extrahieren einen Bruchteil eines Dokuments. Dieser Beitrag beschreibt die Entwicklung eines modularen Systems zum multilingualen Question Answering. Die Strategie bei der Entwicklung zielte auf eine schnellstmögliche Verwendbarkeit eines modularen Systems das auf viele frei verfügbare Ressourcen zugreift. Das System integriert Module zur Erkennung von Eigennamen, zu Indexierung und Retrieval, elektronische Wörterbücher, Online-Übersetzungswerkzeuge sowie Textkorpora zu Trainings- und Testzwecken und implementiert eigene Ansätze zu den Bereichen der Frage- und Antwort-Taxonomien, zum Passagenretrieval und zum Ranking alternativer Antworten.

## **Abstract:**

Question answering systems try to find an answer for a specific question. In order to do that, these systems search within a document collection and extract a fraction of a document. This paper describes the development of a modular multi-lingual question answering system. The system was developed by applying a fast-forward approach using many existing resources. The system integrates named entity recognition, indexing, retrieval, electronic terminologies, machine translation and corpora for training and testing and implements question and answering taxonomies, passage retrieval and ranking for answer candidates.

## **1 Einleitung**

Die theoretische Grundlagen zum Question Answering reichen weit zurück (Lehnert 1978, Nebel & Marburger 1982). In den vergangenen Jahren hat die

Thematik stark an Dynamik gewonnen und besonders das sog. Open-Domain Question Answering hat die Entwicklungen im Bereich der maschinellen Sprachverarbeitung nachhaltig beeinflusst.

Ein Open-Domain Question Answering System akzeptiert natürlichsprachliche Anfrage und liefert dazu eine passende Antwort. Beispielsweise sollte auf die Frage „Wer war Helmut Kohl?“ die Antwort „deutscher Bundeskanzler“ geliefert werden. Question Answering steht somit in mehreren Traditionen. Während das Information Retrieval sich bis heute eher als Dokument Retrieval präsentiert und als Antwort lediglich vollständige Dokumente liefert, leistet das Question Answering einen weiteren Schritt und extrahiert die möglicherweise kurze Antwort aus dem Dokument. Dazu bedient sich das Question Answering im ersten Schritt meist der Verfahren des Information Retrieval zur Identifikation der relevanten Dokumente. Anschließend kommen häufig computer-linguistische Verfahren zur Extraktion der Antwort zum Einsatz.

Während Question Answering sich also vor allem der robusten und statistischen Verfahren des Information Retrieval bedient, erinnert das Anwendungsszenario eher an Visionen der Künstlichen Intelligenz, in denen es darum ging, das Alltagswissen zu erfassen und so beliebige Fragen beantworten zu können (??). Dabei kommen jedoch völlig andere Verfahren zum Einsatz, die sich für Masendaten nicht so gut eignen.

Die andauernde Aktualität dieser Thematik hat dazu geführt, dass Question Answering mittlerweile ein fester Bestandteil der Evaluierungsinitiativen von TREC, CLEF und NTCIR<sup>1</sup> ist. Die dort verwendeten Methoden sowie die erreichten Ergebnisse (mit einer durchschnittlichen Precision von 20%) können als repräsentativ angesehen werden. Die Technologie für das Question Answering ist demnach noch nicht völlig ausgereift.

Das sprachübergreifende Question Answering erhöht die Komplexität noch mehr. Liegt die Antwort nicht in der gleichen Sprache vor wie die Frage, so erfordert dies sprachübergreifende Technologie. Die Informationswissenschaft der Universität Hildesheim greift auf Erfahrungen zum sprachübergreifenden Information Retrieval (cf. Hackl et al. 2004) und zur Eigennamen-Erkennung (Mandl et al. 2005) zurück. Der Beitrag beschreibt im Folgenden die Systemarchitektur, einzelne Systemmodule und berichtet über erste Ergebnisse bei der Evaluierung im Rahmen der europäischen Initiative Cross Language Evaluation Forum (CLEF).

Das hier beschriebene System wurde im Wintersemester 2004/05 im Rahmen eines Projektseminars im Rahmen der Studiengänge *Internationales Informationsmanagement*<sup>2</sup> (IIM) und *Informationsmanagement/Informationstechnologie*<sup>3</sup> (IMIT) an der Universität Hildesheim entwickelt (Bach et al. 2005). Pro-

---

<sup>1</sup> <http://trec.nist.gov>; <http://www.clef-campaign.org>, <http://research.nii.ac.jp/ntcir>

<sup>2</sup> <http://www.uni-hildesheim.de/de/5806.htm>

<sup>3</sup> <http://www.uni-hildesheim.de/de/imit.htm>

jekturse im Hauptstudium bearbeiten größere Aufgaben um so die konstruktiven Fähigkeiten bei der Erstellung von realitätsnah Informationssystemen zu schulen .

## 2 Question Answering System

Das bislang erstellte System verarbeitet in einem einzigen Durchlauf ohne rekursive Optimierung einzelne Fragen und liefert die vom System gefundenen Antworten in den Textkorpora der Sprachen, die zu Beginn durch den Benutzer spezifiziert wurden. Zum gegenwärtigen Stand erlaubt das System Question Answering für die Sprachpaare deutsch-deutsch, deutsch-englisch, englisch-deutsch und englisch-englisch.

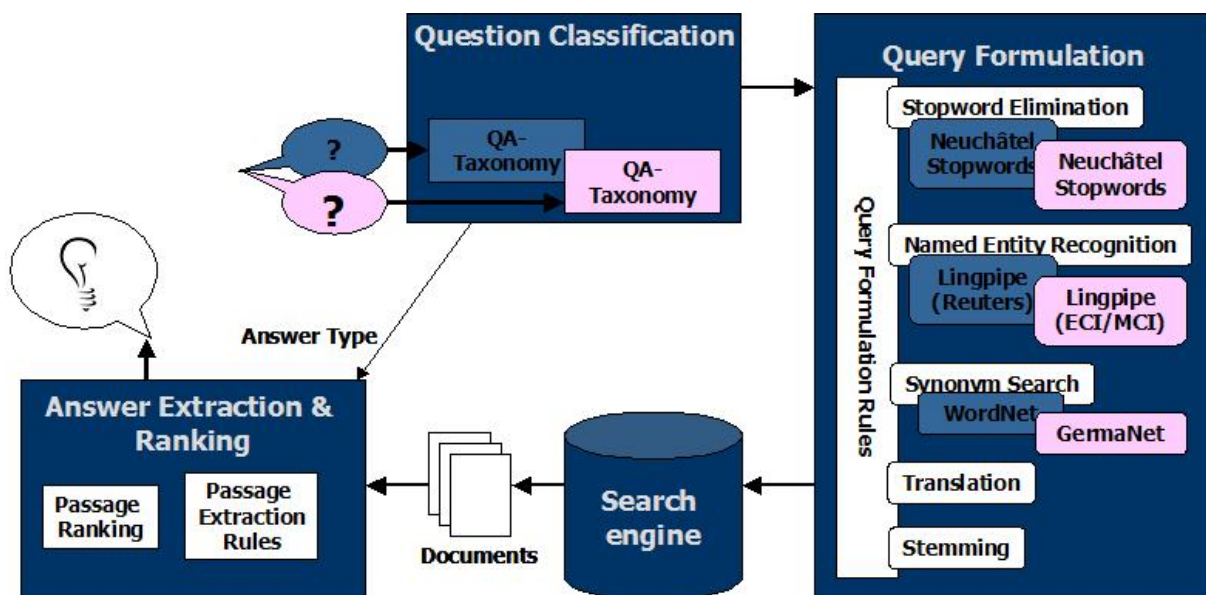


Abbildung 1: Systemüberblick

### 2.1 Systemarchitektur

Die Systemarchitektur folgt im Grossen und Ganzen der Architektur, die sich in den vergangenen Jahren innerhalb der Forschungsgemeinschaft durchgesetzt hat und besteht aus den nachstehenden Hauptkomponenten:

- Frageanalyse entsprechend einer eigens entwickelten Fragen-Taxonomie,
- Frageverarbeitung bestehend aus
  - Stoppworteliminierung,
  - Eigennamenerkennung,
  - Schlüsselwortextraktion sowie Schlüsselwortexpansion
- Übersetzung der Frage, je nach Maßgabe durch den Benutzer
- Retrieval- und Extraktionskomponente zum
  - Dokumentretrieval relevanter Dokumente entsprechend der generierten Schlüsselwörter

- Passagenretrieval entsprechend einer Formel zur Berechnung von Schlüsselwortauffreten und –dichte.
- Ranking der extrahierten Passagen entsprechend a priori erstellter Regeln, die mit dem jeweiligen Frage- und daraus folgenden Antworttyp korrespondieren.

## 2.2 Frage- und Antwort-Taxonomie

Aufgrund der Einbindung der Forschungsaktivitäten am Arbeitsbereich Informationswissenschaft der Universität Hildesheim in die CLEF-Evaluierungsinitiative wurde die Erstellung der Frage- und Antwort-Taxonomie an den Fragen der CLEF-Question-Answering Tracks der Jahre 2003 und 2004 (Magnini et al. 2004, Magnini et al. 2005) und einem Erfahrungsbericht der Gruppe um Moldovan und Harabagiu (1999) ausgerichtet.

Entsprechend den Vorgaben für die Erstellung des Systems besteht die Hauptaufgabe dieses Arbeitspakets in der Erstellung einer Frage-Taxonomie für faktische Fragen mit einfacher Semantik, vor allen Dingen um die charakteristischen W-Fragen wie „Wer, Was, Wann, Wo, Wie?, usw.“, sowie der Zuweisung eines oder mehrerer korrespondierender Antworttypen, die die Minimierung bzw. das Ranking der Antworten aus dem Dokumentkorpus minimiert.

Das System überprüft die ihm vorgelegten Antworten in zwei Schritten. Zunächst wird nach dem Vorkommen eines Frageworts gesucht, in einem zweiten Schritt nach Signalwörtern, die auf einen bzw. den speziellen Fragetyp schließen lassen, sofern eine Subspezifikation nötig ist. Wird weder Frage- noch Signalwort gefunden, wird kein Antworttyp generiert.

Zur Illustration dieser Vorgehensweise sei als Beispiel die Frage: “In **welcher Stadt** wurde Emil Zapotek geboren?” gegeben. Der Fragetyp, der hierbei ermittelt wird, ist “welcher”, der Antworttyp ist jedoch nur anhand des nachstehenden Worts “Stadt“ zu ermitteln und ergibt Location, also einen bestimmten Ort, nachdem gesucht wird.

Im Gegensatz zur Frage-Taxonomie, die möglichst fein gehalten wurde, ging das Bemühen bei Erstellung der Antwort-Taxonomie (siehe Abb. 2) dahin, diese eher allgemein zu halten, um möglichst viele Fragen abzudecken.

Nach Erstellung der beiden Taxonomien wurde diese anhand der Fragen aus den Question Answering Tracks in CLEF aus den Jahren 2003 und 2004 überprüft. Das Gesamtergebnis sieht so aus, dass 73% der CLEF-Fragen von 2004 einer korrekten Antwortkategorie zugewiesen wurden, bei weiteren 14% der Fragen wurde die Richtung der Antwortkategorie richtig identifiziert und 14% der Fragen wurde falsch bestimmt. Dies entspricht im Fall des 200 Fragen umfassenden Fragenkorpus aus CLEF 2004 einer Quote von 86% korrekt identifizierter Fragen. Am zuverlässigsten bestimmt das System Fragen, die *Wann*, *Womit* und *Wozu* enthalten. Dabei gab es niemals eine falsche Bestimmung des Antworttyps. Weiterhin sehr zuverlässig ist die Bestimmung von *Wo*-Fragen und

die Aufforderung irgendetwas zu *Nennen*. Fragen, die mit *Wer* oder *Wie* beginnen werden gut, *Was*-Fragen eher ungenügend gut bestimmt.

Daraus wird ersichtlich, dass Fragen, die Unterklassen zu ihrer Bestimmung benötigen, weniger präzise erkannt werden, als Fragen, die nach der ersten Stufe sofort eingeordnet werden können.

Antwortkategorie		Fragewort (+ Erweiterung)
Alter	Quantitativ	Wie alt?
Anzahl	Quantitativ	Wie viele?
Art und Weise	Prozedural	Wie?
Datum	Temporal	Wann?
Eigenname Person/Firma		Wer ist + der/die/das Wem? Wen? Welcher? Welche? Welches? Welchen? Welchem? Wie + lautet? Wie + heißt? Wie ist (der Name)?
Entfernung	Quantitativ	Wie lang? Wie weit?
Grund	Kausal	Warum? Wodurch? Wozu?
Liste		Nenne Nennen Welche + [...]?
Maßeinheit/Währung	Quantitativ	Wie viel?
Material, Teile eines Ganzen		Woraus?
Mittel	Instrumental	Womit?
Ort	Lokativ	Wo? Woher? Wohin?
Zeitangabe mit Einheit	Temporal	Wie + lange?
Zweck	Final	Wozu?

Abbildung 2: Antwort-Taxonomie

### 2.3 Stoppworteliminierung

Für die Stoppworteliminierung wurden die Stoppwortlisten<sup>4</sup> der Universität Nêuchatel für Englisch und Deutsch integriert. Anhand dieser Positiv-Liste wurden alle Vorkommen dieser Wörter aus den Fragen entfernt, um diese daraufhin zu überprüfen, ob es sich bei einem oder mehreren um Eigennamen handelt und ob sich für die verbleibenden Wortformen Synonyme zur Expansion der Schlüsselwörter finden lassen. Beide Arbeitsschritte sollen im Folgenden genauer erläutert werden.

<sup>4</sup> <http://www.unine.ch/info/clef/>

## 2.4 Eigennamenerkennung

Zur Eigennamenerkennung wurde auf das für Forschungszwecke frei verfügbare System LingPipe<sup>5</sup> zurück gegriffen. Lingpipe verwendet einen statistischen Ansatz zum maschinellen Lernen von Eigennamen und Kategorien, die anhand von Korpora trainiert werden, welche vorab mit Labels versehen wurden. In einer früheren Arbeit wurde LingPipe als bestes Werkzeug zur Eigennamenerkennung evaluiert (Mandl et al. 2005). Es entstanden somit zwei Arbeitspakete:

- a) für das Training anhand geeigneter Korpora
- b) für die Anwendung der Trainingsdaten bzw. der Anpassung des Systems an das Gesamtsystem zum Question Answering. Dabei sollten sowohl in den Fragen als auch in den Textkorpora, in denen die Antworten zu finden waren, die Eigennamen erkannt und markiert werden.

Für das Deutsche wurde LingPipe anhand des Linguistic Data Consortium (LDC<sup>6</sup>) Korpus der Frankfurter Rundschau trainiert, dessen Umfang ca. 36 Millionen fortlaufende Wortformen entspricht, für das Englische wurde das English Language News von Reuters verwendet, das 810.000 Nachrichtentexte beinhaltet, verwendet. Eine Evaluation der Erkennungsrate ergab einen Wert von 60% für die korrekt erkannten Eigennamen sowie einen Wert von 42 % für die korrekte Kategorisierung. LingPipe kategorisiert die erkannten Eigennamen in vier verschiedene Gruppen: Person (PER), Organisation (ORG), Ort (LOC) und Sonstiges (MISC).

Anschließend wurde das trainierte System an das Question Answering System so eingebunden, dass aus den Fragen die jeweiligen Eigennamen extrahiert und aus den Korpora, auf denen die Fragen gesucht wurden, eine Eigennamenindexierung erstellt wurde.

## 2.5 Synonymsuche

Zur Verbesserung des Retrievalvorgangs wurde wie in vielen anderen Systemen zur Expansion der Schlüsselwörter die Synonymiefunktion von WordNet<sup>7</sup> (Fellbaum 1996) verwendet. Da es sich um ein cross-linguales System handelt wurden für deutsche Anfragen auf GermaNet<sup>8</sup> (Hamp & Feldwig 1997) zurückgegriffen. WordNet ist ein monolinguales Lexikon für das Englische, das sich trotz einiger Schwächen zu einem Quasi-Standard für eine Vielzahl von Anwendungen der maschinellen Sprachverarbeitung entwickelt hat. In beiden Fällen wurden die jeweils aktuellen Versionen lokal installiert und in das System eingebunden.

---

<sup>5</sup> <http://www.alias-i.com/lingpipe/>

<sup>6</sup> <http://www ldc.upenn.edu/>

<sup>7</sup> <http://wordnet.princeton.edu/doc>

<sup>8</sup> <http://www.sfs.nphil.uni-tuebingen.de/lstd/>

## 2.6 Übersetzung

Die Übersetzungsanteile wurden über Anfragen an Babylon, einer im Rahmen eines früheren Projektseminars entwickelten Java-Anwendung zur Meta-Online-Übersetzung bewerkstelligt (Krauss & Petzold 2004). Dabei werden die Anfragen von Babylon an vier freie Übersetzungsanbieter im Internet gesendet und an das Question Answering-System zurück geliefert. Anschließend werden von Systemseite die relevanten Ergebnisse der Übersetzung als Vereinigungsmenge dargestellt.

## 2.7 Passagenretrieval

Die überwiegende Mehrheit der Systeme zum Question Answering unterscheidet sich von herkömmlichen Retrievalverfahren durch das sog. Passagenretrieval (im engl. Passage-Retrieval). In Ansätzen, die sich stärker dem Text Mining verbunden fühlen, wird von daher Question Answering häufig auch als Answer Mining bezeichnet, ohne dass sich die dabei verwendeten Verfahren grundlegend voneinander unterscheiden. In sämtlichen Fällen werden Retrieval- und Mining-Verfahren auf Textpassagen der unterschiedlichsten Länge angewendet.

Im vorliegenden Fall wurde ein gemischter Ansatz aus Text- und Passagenretrieval unter Verwendung der open-source Suchmaschine Lucene<sup>9</sup> sowie eines quantitativ arbeitenden Ansatzes zum Passagenretrieval entwickelt. Dies bedingt, dass nach Indexierung der Dokumente zunächst ein herkömmliches Dokumentenretrieval durchgeführt wurde, welche die aus der Frage verbliebenen Schlüsselwörter mit den Dokumenten vergleicht. Auf den ausgewählten Dokumenten wurde mittels nachstehender Formel anschließend ein Passagenretrieval auf Passagen einer Länge von 200 Zeichen plus den darauf folgenden Zeichen bis zum nächsten Punkt durchgeführt. Das Verfahren zum Passagenretrieval ist eng an den von MITRE (Light et al. 2001) entwickelten Algorithmus angelehnt und berechnet für jede Passage eines Dokuments, in dem ein oder mehrere Schlüsselwörter auftreten, einen Wert, der sich aus Auftreten und Anzahl aller Schlüsselwörter berechnet.

$$\text{Formel 1: } \text{Score} = \underbrace{\sum_{t=0}^n (k_1 \times b_t \times \frac{1}{A_t})}_{\text{Relevanzwert 1}} + \underbrace{k_2 \times \frac{f_{p,t}}{A_p}}_{\text{Relevanzwert 2}}$$

Relevanzwert 1 berechnet sich aus dem booleschen Wert  $b_t$  für die einzelnen Terme  $t$ . Dieser Wert hat dementsprechend entweder die Ausprägung 0 für kein Auftreten sowie 1 für ein oder mehrere Auftreten von  $t$ . Dieser Wert wird mit  $A_t$ ,

---

<sup>9</sup> <http://lucene.apache.org/>

d.h. der absoluten Anzahl der Frageterme  $t$  sowie einer heuristisch ermittelten Konstante  $k_1$  berechnet.

Relevanzwert 2 trägt der Häufigkeit der einzelnen Terme  $t$  in der Passage  $p$  Rechnung und ergibt sich aus dem Quotienten  $f_{p,t}$ , d.h. der Häufigkeit von  $t$  in Passage  $p$  und der Anzahl der Wörter in Passage  $p$ , der wiederum mit einer heuristischen Konstante  $k_2$  multipliziert wird. Die Summe beider Relevanzwerte wird über sämtliche Terme der Suchanfrage aufsummiert.

Dieses Verfahren wurde aufgrund folgender Überlegungen entwickelt: So soll das mehrmalige Auftreten eines Terms nicht so stark gewichtet werden wie das Auftreten mehrerer verschiedener Terme. Dieser Anforderung wird Relevanzwert 1 gerecht. Um ein häufiges Auftreten eines Suchterms jedoch nicht gänzlich unberücksichtigt zu lassen, wird für jedes Auftreten eines Suchterms ein von der Anzahl der Wörter in der Passage abhängiger Wert addiert. Das Verfahren berücksichtigt bislang weder die Synonyme noch die Dichte der Schlüsselwörter. Die ersten Ergebnisse zeigen allerdings, dass der Algorithmus ausreichend gute Ergebnisse liefert.

## 2.8 Antwort-Ranking

Im Gegensatz zu herkömmlichen Verfahren des Information Retrieval bestehen die Anforderungen des Question Answering darin, eine, d.h. genau eine Antwort auf eine (Suchan-)Frage zu liefern, die nicht aus einem Dokument, sondern aus einer Textsequenz oder idealerweise aus einer korrekt formulierten Antwort besteht. Die Menge der im Passage-Retrieval gelieferten Antworten, müssen demnach auf ein oder kein Element reduziert werden. Letzterer Fall wäre dann gegeben, wenn die zur Verfügung stehende Textmenge die Antwort auf die gegebene Frage nicht enthält.

Die Auswahl bzw. Bewertung der durch das Passagenretrieval gelieferten Antworten wird demnach von zwei Faktoren bestimmt: a) dem maximalen Wert des Passagenretrievals, b) einer Menge von Regeln, welche die Antworten mit der Vorgabe der Antwort-Taxonomie vergleicht (siehe Abschnitt 2.2). So sinkt die Wahrscheinlichkeit, dass es sich um eine korrekte Antwort handelt trotz des höchsten Werts für das Auftreten der Schlüsselwörter, wenn bspw. kein Wort der Antwort mit dem Antworttyp korrespondiert, bspw. wenn nach einer Stadt gesucht wird, die Antwort jedoch keinen Städtenamen enthält. In diesem Fall müsste einer weiteren Antwort, die zwar einen niedrigeren Wert, dafür jedoch einen Ortsnamen enthält, der Vorzug gegeben werden.

Aus diesem Grund wurden nach obigem Muster eine Reihe von Regeln spezifiziert, welche die hoch bewerteten Passagen noch einmal auf ihre Verwertbarkeit überprüfen. Nach Abarbeitung dieser Regeln wird dann der verbleibende optimale Antwortkandidat vom System als Antwort auf die zu bewertende Frage ausgegeben, andernfalls (d.h. wenn kein Antwortkandidat verblieb) wurde als Antwort die Zeichenkette NIL (für Not In List) ausgegeben.



## 2.9 Heuristiken

Zusätzlich wurden einige plausible Heuristiken eingebaut, die vor allem die Treffergenauigkeit für Anfragen zu Eigennamen erhöhen sollen.

- Wird ein Eigenname erwartet und die Antwortmenge enthält Dokumente mit Eigennamen, so werden nur diese zum Passagen-Retrieval weitergeleitet.
- Wird ein Eigenname erwartet, dann werden die häufigsten Eigennamen des korrekten Typs in der Antwortmenge bestimmt und jeweils die erste Passage mit diesen Eigennamen als Antwort geliefert.
- Wird ein Eigenname erwartet und keine Antwort mit einem Eigennamen gefunden, so liefert das System die ersten 90 Zeichen der Passage an erster Stelle in der Liste von Lucene.
- Triviale Antworten werden nicht ausgegeben. Antworten gelten dann als trivial, wenn sie nur aus einem Wort bestehen, aus dem Namen einer Nachrichtengagentur bestehen, oder eine Untermenge der Frage bilden.
- Als Konfidenzwert wird der Lucene-Ranking-Wert der Antwort-Passage geliefert.

## 3 Evaluierung und Ergebnisse

Während der Entwicklung wurde das aus vielen Komponenten bestehende System fortlaufend auf Stabilität evaluiert. Für eine heuristische Evaluierung der Effektivität der Beantwortung wurde eine Benutzungsoberfläche erstellt, die Abbildung 3 zeigt. Für die Evaluierung steht zusätzlich eine Kommandozeilen-Variante der Anwendung zur Verfügung, um im Batch-Betrieb automatisiert eine Vielzahl von Anfragen an das System zu übergeben. Damit konnte das System im Rahmen der Evaluierungsinitiative CLEF im Jahr 2005 evaluiert werden (Strötgen et al. 2005). Wie Tabelle 1 zeigt, war die Qualität der Ergebnisse lediglich für Definitionsfragen zufriedenstellend. Das System konnte 10 Fragen korrekt beantworten, 29 wurden als 29 Systemantworten wurden von den Juroren als ungenau eingestuft und 161 als falsch.

**Table 1.** Results for QA system of the University of Hildesheim in 2005

<b>Sprachen</b>	<b>Question Type</b>	<b>Accuracy</b>
English -> Deutsch	Definition	18,00%
English -> Deutsch	Fakten	0,83%
English -> Deutsch	alle	5,00%

Die mangelhafte Performanz zeigt, dass das System vor der Teilnahme an CLEF nicht evaluiert wurde. Ein Nachteil des Systems liegt in der starren Länge der Passagen von 200 Zeichen. Sehr kurze Antworten wurden daher nicht korrekt erkannt.

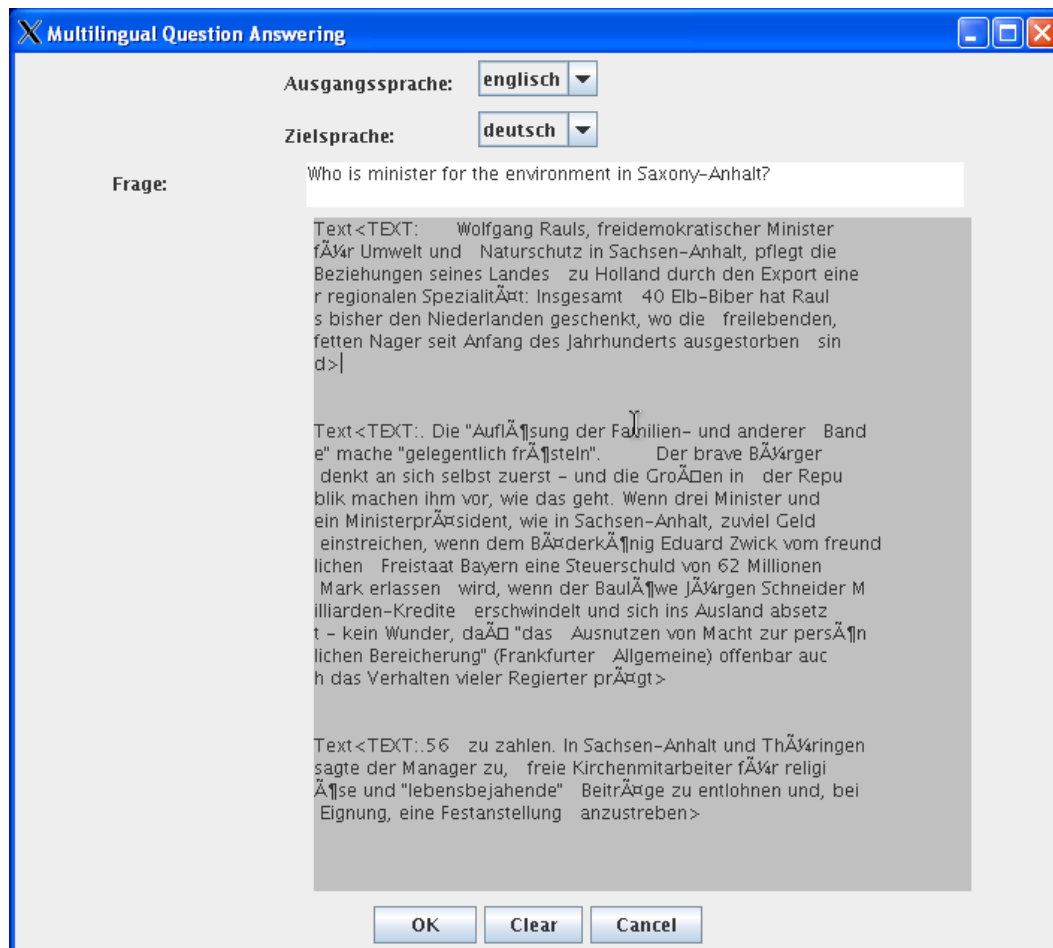


Abbildung 34: Eingabemaske

## 4 Schlussbemerkungen

Robuste und effektive Systeme zum multilinguales Question Answering im Internet können einen erheblichen Gewinn für viele Benutzer mit spezifischen Informationsbedürfnissen darstellen. Wichtig wird für solche Systeme die Auswahl einer vertrauenswürdigen Quelle.

<sup>10</sup> <http://bscw.fit.fraunhofer.de/>

<sup>11</sup> <http://www.eclipse.org/>

<sup>12</sup> <http://subversion.tigris.org/>

<sup>13</sup> <http://subclipse.tigris.org/>

<sup>14</sup> <https://www.cvshome.org/>

## 5 Literatur

- Bach, Kerstin; Marco Fischer; Tanja Mackensen; Ulrike Maichel; Curt Nowak; Carina Völpel; Kathrin Wünnemann (2005): Projektdokumentation Question Answering. Arbeitsmaterialie. Informationswissenschaft, Universität Hildesheim.
- Fellbaum, Christiane (1996): WordNet: Ein semantisches Netz als Bedeutungstheorie. In: Grabowski, Joachim; Herrmann, T.; Harras, G. (Hrsg): *Bedeutung, Konzepte, Bedeutungskonzepte*, Opladen, Westdeutscher Verlag, 1996, pp. 211 - 230.
- Hackl, René; Kölle, Ralph; Mandl, Thomas; Ploedt, Alexandra; Scheufen, Jan-Hendrik; Womser-Hacker, Christa (2004): Multilingual Retrieval Experiments with MIMOR at the University of Hildesheim. In: Peters, C.; Braschler, M.; Gonzalo, J.; Kluck, M. (Hrsg.): *Comparative Evaluation of Multilingual Information Access Systems: 4<sup>th</sup> Workshop of the Cross-Language Evaluation Forum, CLEF 2003, Revised Selected Papers*. Trondheim, Norway, August 21-22, Springer [LNCS 3237]
- Hamp & Feldwig 1997??
- Harabagiu, Sanda; Moldovan, Dan (2003): Question Answering. In: Mitkov, Ruslan: *The Oxford Handbook of Computational Linguistics*. Oxford, University Press.
- Krauss, Sebastian; Petzold, Jörg (2004): Babylon: Implementierung von Wrappern für Übersetzungsdienste. Arbeitsmaterialie. Informationswissenschaft, Universität Hildesheim.
- Lehnert, Wendy (1978): The Process of Question Answering: a computer simulation of cognition. Lawrence Erlbaum.
- Light, Marc, Mann, Gideon S., Riloff, Ellen, Breck, Eric (2001): Analyses for elucidating current question answering technology. Journal of Natural Language Engineering, Special Issue on Question Answering Fall-Winter 2001.
- Magnini, Bernardo; Simone Romagnoli, Alessandro Vallin, Jesús Herrera, Anselmo Peñas, Victor Peinado, Felisa Verdejo; Maarten de Rijke (2004): The Multiple Language Question Answering Track at CLEF 2003. In: Peters, C.; Braschler, M.; Gonzalo, J.; Kluck, M. (Hrsg.): *Comparative Evaluation of Multilingual Information Access Systems: 4<sup>th</sup> Workshop of the Cross-Language Evaluation Forum, CLEF 2003, Revised Selected Papers*. Trondheim, Norway, August 21-22, Springer [LNCS 3237]
- Magnini, Bernardo; Alessandro Vallin, Christelle Ayache, Gregor Erbach, Anselmo Peñas, Maarten de Rijke, Paulo Rocha, Kiril Simov and Richard Sutcliffe (2005): Multiple Language Question Answering (QA@CLEF). Overview of the CLEF 2004 Multilingual Question Answering Track. In: *Working Notes 5<sup>th</sup> Workshop of the Cross-Language Evaluation Forum, CLEF 2004*. Bath, England.
- Mandl, Thomas; Schneider, René; Schnetzler, Pia; Womser-Hacker, Christa (2005): Evaluierung von Systemen für die Eigennamenerkennung im cross-lingualen Information Retrieval. In: *Gesellschaft für linguistische Datenverarbeitung. Beiträge der GLDV-Frühjahrstagung*. Bonn, 30.3. - 01.04. Frankfurt a. M. et al. Peter-Lang.
- Strötgen, Robert; Mandl, Thomas; Schneider, René (2005): A Fast Forward Approach to Cross-lingual Question Answering for English and German. In: *Working Notes of the 6<sup>th</sup> Workshop of the Cross-Language Evaluation Forum, CLEF 2005*. Sept. 2005, Wien.
- Moldovan, Dan; Sanda Harabagiu; Marius Paşca; Rada Mihalcea; Richard Goodrum; Roxana Gîrju; Vasile Rus (1999): LASSO: A tool for Surfing the Answer Net. In: *Proceedings of the Eighth Text Retrieval Conference (TREC-8)* S. 175-184.  
<http://trec.nist.gov/pubs/trec8/papers/smu.pdf>

Nebel, Bernhard; Marburger, Heinz (1982): Das natürlichsprachliche System HAM-ANS: Intelligenter Zugriff auf heterogene Wissens- und Datenbanken. *In: GI Jahrestagung 1982, Gesellschaft für Informatik*. S. 392-402.

Voorhees, Ellen (2002): Overview of the TREC 2002 Question Answering Track. *In: Proc 11<sup>th</sup> Text REtrieval Conference (TREC 2002)*.  
<http://trec.nist.gov/pubs/trec11/papers/QA11.pdf>