REALISTISCHE BEWERTUNG VON RETRIEVALVERFAHREN

Ulrich Pfeifer
Universität Dortmund

Realistische Bewertung von Retrievalverfahren

Bericht von der zweiten
TREC-Konferenz
(30.8-2.9.93 in Was hingt on)

Ziele der Initiative

Ziel der US-amerikanischen TREC (Text Retrieval Conference) Initiative ist es, zum einen Standard-Kollektionen von Texten zu erstellen, auf deren Basis verschiedene Retrieval-Ansätze verglichen werden können. Zum anderen sollen Kollektionen einer Größe aufgebaut werden, die realistischen Anwendungen nahe kommen. Damit soll das alte Vorurteil der Untauglichkeit neuerer IR-Methoden (bzgl. Effizienz und Effektivität) für große IR-Datenbasen widerlegt werden.

Inhalt

Im Rahmen der Initiative wurde eine Kollektion von mittlerweile 3 GB Text (Zeitungsmeldungen, Patentschriften, Referate wissenschaftlicher Artikel, ...) aufgebaut, die auf drei CD-ROMs an die Teilnehmer verteilt wurden.

Die Teilnehmer hatten zwei Arten von Aufgaben zu lösen:

ad-hoc: Eine Liste von 50 Fragen im Umfang von je etwa einer A4-Seite Text, die gegen eine Teilkollektion (1 bzw. 2 GB) von Dokumenten getestet werden sollten.

routing: Weitere 50 Fragen, für die für eine CD Feedback-Daten zur Verfügung standen, und die dann für eine weitere CD prozessiert werden sollten.

Die Teilnehmer (etwa 30 Systeme) sandten für jede Frage eine Rangliste von Dokumenten als Antwort ein, für die dann Relevanzurteile erstellt wurden. Darauf aufbauend konnte die Retrievalqualität der Systeme verglichen werden.

Kategorien für die Teilnahme

Die Teilnehmer konnten verschiedene Kategorien von Ansätzen verfolgen:

1. Vollautomatische (initiale) Frage-Erstellung
2. Manuelle (initiale) Frage-Erstellung
3. Manuelle (initiale) Frage-Erstellung mit Feedback. *Das heißt eigenes Feedback bei den ad-hoc Fragen*

Weiter wurde zwischen voller (Kategorie A) und eingeschränkter Teilnahme (Kategorie B) unterschieden. Nur Kategorie-A-Teilnehmer mußten mit der vollen Datenmenge arbeiten.

Ergebnisse

Ohne der genaueren Auswertung vorgreifen zu wollen, lassen sich tendenziell folgende Aussagen machen.

=> Bei den ad-hoc Fragen schneiden voll-automatische Systeme im Mittel besser ab als Systeme, die auf manueller Konstruktion der Fragen beruhen. Andererseits gibt es bei automatischen Systemen meist eine breitere Streuung der Ergebnisse zwischen einzelnen Fragen.

Die sechs besten automatischen Systeme basierten auf dem Vektorraummodell bzw. verschiedenen probabilistischen Modellen. Das System von Cornell, Siemens und Carnegie Mellon basierten auf dem Vektor-raummodell, wobei die beiden letzteren Thesauri zur Frageerweiterung verwendeten. Die Systeme von Berkley, Dortmund und University of Massachusettes basieren auf probabilistischen Modellen, wobei letzteres ein probabilistisches Netzwerk als Retrie-valmaschine verwendet.

Bei den manuellen Systemenschnitten das System der UMASS bzw. Carnegie Mellon am besten ab, die sich auf Korrekturen der automatisch erstellten Fragen beschränkten. N ur diese beiden Systeme konnten sich mit den automatischen Systemen messen. Die rein manuellen Systeme waren signfikant schlechter.

=> Bei den routing Fragen entstand bezüglich des Vergleichs von automatischen mit manuellen Systemen das gleiche Bild. Hier sind die Unterschiede nur noch signifikanter als bei den ad-hoc Fragen.

Bei den automatischen Systemen war das System aus Cornell dem aus Dortmund überlegen, das wiederum signifikant besser als der Rest der Mitbewerber abschnitt.

Neben den Systemen der University of Massachusettes und Carnegie Mellon schnitten die Systeme von General Electrics und TRW gut ab, die konventionelle boolesche Systeme einsetzten.

=> Die von vielen Anbietern kommerzieller Systeme unterstellten Effizienzprobleme gibt es tatsächlich: bei den kommerziellen Systemen! Während man bei einem Firmenprodukt 8 Minuten auf eine Antwort warten muß, liefert beispielsweise das SMART-System (Cornell) auf einer SUN Workstation bei einer Frage (mit 50 ... 100 Wörtern) bezüglich einer 2GB Kollektion die Antwort innerhalb von 10 Sekunden ab.

Ausblick

Die TREC-Initiative wird fortgesetzt und geht damit schon in das dritte Jahr. In den neuen Kollektionen werden wohl zum einen fremdsprachige Teile (als Option, wahrscheinlich Spanisch) zum anderen *noisy*-Teile wie e-mail aufgenommen werden.

Informationen

Die Proceedings zur ersten TREC-Konferenz wurden vom National Institute of Standards and Technology (NIST) als Special Publication No. 500-207 veröffentlicht.

Redaktionsschluß für den zweiten Tagungsband ist im November diesen Jahres, der kurz darauf von gleicher Quelle zu beziehen sein wird.