



In: Hammwöhner, Rainer; Wolff, Christian; Womser-Hacker, Christa (Hg.): Information und Mobilität, Optimierung und Vermeidung von Mobilität durch Information. Proceedings des 8. Internationalen Symposiums für Informationswissenschaft (ISI 2002), Regensburg, 8. – 11. Oktober 2002. Konstanz: UVK Verlagsgesellschaft mbH, 2002. S. 225 – 237

## **Das Cross-Language Evaluation Forum (CLEF) - Evaluationsumgebung und Forschungskontext für mehrspra- chiges Information Retrieval (mit einer Skizze der Ergebnisse von CLEF 2002)**

*Michael Kluck*

Informationszentrum Sozialwissenschaften (IZ), Bonn /  
Humboldt-Universität zu Berlin, Abteilung Pädagogik und Informatik  
michael.kluck@educat.hu-berlin.de

### **Zusammenfassung**

Einleitend werden die Struktur und Methodik von CLEF kurz erläutert. Über ausgewählte Ergebnisse der CLEF-Kampagne 2002 wird berichtet, insbesondere über die verwendeten Verfahren und Ressourcen für die Behandlung der Multilingualität sowie über das interaktive und das fachbezogene Cross-Language Information Retrieval (CLIR). Die Möglichkeiten zur Beteiligung an der Forschung im Bereich des mehrsprachigen Information Retrieval (CLIR) und der Methodologie der IR-Evaluationen werden aufgezeigt. Abschließend wird die weitere Entwicklung von CLEF skizziert.

### **Abstract**

At the beginning structure and methodology of CLEF are briefly explained. Selected results of the CLEF 2002 campaign will be reported, especially the used procedures and resources for the treatment of multilinguality and the interactive and scientific cross-language information retrieval (CLIR). The possible fields of research in the area of CLIR and its evaluation methodology are shown. Finally the further development of CLEF is shortly described.

## **1 Einleitung: CLEF im Kontext der IR-Forschung**

Das Cross-Language Evaluation Forum (CLEF) ist als ein Spin-off der TREC-Evaluierungskampagne aus dem damaligen TREC-Task Cross-Language Information Retrieval (CLIR) entstanden (näheres siehe Harman et al. 2001 und Kluck et al. 2002), das sich schwerpunktmäßig mit dem CLIR europäischer Sprachen befasst. Parallel dazu hat sich auch eine entsprechende



Dieses Dokument wird unter folgender [creative commons](http://creativecommons.org/licenses/by-nc-nd/2.0/de/) Lizenz veröffentlicht:  
<http://creativecommons.org/licenses/by-nc-nd/2.0/de/>

Initiative (NTCIR)<sup>1</sup> in Japan entwickelt, die sich auf die fernöstlichen Sprachen (Japanisch, Chinesisch, Koreanisch etc.) konzentriert, während sich TREC nun mit den arabischen Sprachen befasst. Alle drei Initiativen fußen auf der bei TREC weiter entwickelten Cranfield-Methodologie.

Die Anforderungen an das mehrsprachige Information Retrieval (CLIR) gehen über das des einsprachigen IR insofern hinaus, als die Transponierung der Anfrage aus einer Sprache in die weiteren Sprachen der durchsuchten Dokumentmengen erforderlich ist und eine integrierte gerankte Antwortliste mit verschiedensprachigen Ergebnisdokumenten gefordert wird.

## **2 CLEF – Struktur und Methoden der Evaluation**

### **2.1 Verwendete Datenbestände und Sprachen der Dokumente**

Die Datenbestände, die in CLEF Verwendung finden, umfassen vorwiegend Zeitungsartikel und Meldungen von Nachrichtenagenturen. Jeder Teilbestand in den verschiedenen Sprachen liegt für das Jahr 1994 vollständig vor. Dadurch soll sichergestellt werden, dass die gleichen Themenkomplexe bzw. Ereignisse in den verschiedenen Sprachen behandelt werden. Die Kernsprachen sind Deutsch Englisch, Französisch, Italienisch und Spanisch. Ferner stehen Datenbestände in Finnisch, Holländisch und Schwedisch zur Verfügung. Für das fachbezogene CLIR stehen Datenbestände aus Fachdatenbanken in Französisch und Deutsch zur Verfügung.

Die folgenden Korpora stehen zur Verfügung:

- Zeitungen und Nachrichtenagenturen in den Kernsprachen
- Englisch (Los Angeles Times 1994): 113 005 Dokumente, 425 MB
- Deutsch (Schweizerische Depeschenagentur – SDA 1994, Der Spiegel 1994-1995, Frankfurter Rundschau 1994): 225 371 Dokumente, 527 MB
- Französisch (Schweizerische Depeschenagentur – SDA 1994, Le Monde 1994): 87 191 Dokumente, 243 MB
- Italienisch (Schweizerische Depeschenagentur – SDA 1994, La Stampa 1994): 108 578 Dokumente, 278 MB
- Spanisch (Agencia EFE 1994): 215 738 Dokumente, 509 MB

---

<sup>1</sup> NTCIR (NII-NACSIS Test Collection for IR Systems); NII = National Institute of Informatics, Tokyo, Japan; NACSIS = National Center for Science Information Systems

wissenschaftliche und domänenspezifische Daten

- Wissenschaften (alle Fachgebiete): AMARYLLIS (Französisch): 148 688 Dokumente, 195 MB
- Sozialwissenschaften: GIRT (Deutsch): 76 128 Dokumente, 150 MB

weitere Daten für bilinguale Tests (Zeitungen und Nachrichtenagenturen)

- Holländisch (NRC Handelsblad 1994-1995, Algemeen Dagblad 1994-1995): 190 604 Dokumente, 540 MB
- Finnisch (Aamulehti Ende-1994-1995) 55 344 Dokumente, 137 MB
- Schwedisch: (Tidningarnas Telegrambyrå 1994-1995) 142 819 Dokumente, 352 MB

Diese Dokumentsammlungen umfassen in der Regel das Jahr 1994 und teilweise auch 1995. Die einzelnen Dokumente der Korpora sind mit SGML-Tags für die Auszeichnung der enthaltenen Datenelemente versehen.

## **2.2 Themenstellungen (Topics)**

Die Erarbeitung der Themenstellungen (Topics) für CLEF ist ein komplexer Prozess, an dem die verschiedenen CLEF-Sprachgruppen (DE, EN, ES, FR, IT, NL, FI, SV) intensiv beteiligt sind. Zusätzlich werden die Themenstellungen von weiteren Gruppen in andere (auch außereuropäische) Sprachen übersetzt (ZH, JP etc.). Die Sprachgruppen müssen Themenstellungen entwickeln, die dem Inhalt der Dokumente entsprechen, die von den verschiedenen Zeitungen und Nachrichtenagenturen im Jahr 1994/1995 gemeldet wurden: z.B. Ereignisse aus Politik, Kultur, Sport, Wissenschaft, Medizin. Jede Sprachgruppe schlägt bis zu 15 Themen vor. Aus diesen mehr als 80 Themenvorschlägen werden schließlich 50 Themenstellungen ausgewählt, die dann als Topics den Teilnehmern der Evaluationskampagne zur Verfügung gestellt werden. Dann werden die 50 Topics aus der jeweiligen Originalsprache in die übrigen Sprachen übersetzt. Schließlich wird von einer unabhängigen Gruppe von Fachübersetzern eine Nachprüfung der Übersetzungen in alle offiziellen Sprachen vorgenommen (topic check), um die Konsistenz und Genauigkeit der Übersetzungen zu sichern.

Für die wissenschaftlichen bzw. domänenspezifischen Aufgabenstellungen im Rahmen von GIRT<sup>2</sup> (German Indexing and Retrieval Database) und AMARYLLIS werden jeweils 25 fachspezifische Themen auf Deutsch bzw. Französisch entwickelt. Für diese wird auch eine englische (und bei GIRT eine

---

<sup>2</sup> GIRT stellt ein Korpus sozialwissenschaftlicher Fachtexte zur Verfügung.

russische Übersetzung) vorbereitet, so dass auch hier CLIR Ansätze getestet werden können.

### 2.3 Topic Beispiele jeweils mit einer Übersetzung

Die Topics sind mit SGML-Tag versehen und bestehen aus einer fortlaufenden Nummer, dem Titel (zwei bis drei Worte zur Kurzcharakterisierung des Themas), der Beschreibung (ein ganzer Satz zur Bezeichnung des Themas) und der ausführlichen Beschreibung (mit Erläuterungen, welche Inhalte einschlägig sind und welche möglicherweise ausgeschlossen werden sollen).

Beispiel 1: Rinderwahnsinn

```
<top>
<num> C088 </num>
<EN-title> Mad Cow in Europe </EN-title>
<EN-desc> Find documents that cite cases of Bovine Spongiform
Encephalopathy (the mad cow disease) in Europe. </EN-desc>
<EN-narr> Relevant documents will report statistics and/or figures on cases
of animals infected with Bovine Spongiform Encephalopathy (BSE),
commonly known as the mad cow disease, in Europe. Documents that only
discuss the possible transmission of the disease to humans are not considered
relevant. </EN-narr>
</top>
<top>
<num> C088 </num>
<IT-title> Mucca pazza in Europa </IT-title>
<IT-desc> Trova i documenti che citano i casi di mucca pazza (Encefalopatia
Spongiforme Bovina) in Europa. </IT-desc>
<IT-narr> Sono rilevanti i documenti che riportano statistiche e/o dati
numerici sui casi di animali affetti da Encefalopatia Spongiforme Bovina
(BSE), comunemente detta morbo della mucca pazza, in tutti i paesi europei.
Non sono rilevanti i documenti sulla possibile trasmissione del morbo
all'uomo. </IT-narr>
</top>
```

### Beispiel 2: Mausarm

```
<top>
<num> C064 </num>
<NL-title> Muisarm </NL-title>
<NL-desc> Zoek documenten waarin melding wordt gemaakt van een
muisarm. </NL-desc>
<NL-narr> Relevante documenten melden klachten die veroorzaakt worden
door het langdurig gebruik van een computermuis. Documenten waarin
maatregelen worden genoemd voor het voorkomen van een muisarm tijdens
het werken achter de computer zijn ook relevant. </NL-narr>
</top>
<top>
<num> C064 </num>
<ES-title> Síndrome RSI y ratones de ordenador </ES-title>
<ES-desc> Encontrar documentos que informen sobre RSI ("repetitive strain
injuries" o "enfermedad del periodista") producidas por el uso del ratón del
ordenador. </ES-desc>
<ES-narr> Los documentos relevantes informan sobre daños causados por el
uso continuado de un ratón de ordenador. Los documentos que proponen
formas de evitar el RSI cuando se usa el ordenador también son relevantes.
</ES-narr>
</top>
```

### Beispiel 3: Eurofighter

```
<top>
<num> C093 </num>
<DE-title> Eurofighter </DE-title>
<DE-desc> Finde Dokumente, die über das Projekt "Jäger 90" bzw. den
"Eurofighter" berichten. </DE-desc>
<DE-narr> Relevante Dokumente berichten über das Projekt "Jäger 90". Die
Partnerländer Deutschland, Großbritannien, Italien und Spanien hatten das
"Eurofighter"-Konsortium gebildet. Von Interesse sind auch der Arbeitsanteil
der beteiligten Rüstungskonzerne sowie Kostenschätzungen. </DE-narr>
</top>
<top>
<num> C093 </num>
<ES-title> Eurofighter </ES-title>
```

<ES-desc> Encontrar documentos que informen acerca del proyecto EFA o "Eurofighter". </ES-desc>

<ES-narr> Los documentos relevantes informan sobre el proyecto EFA. Los países socios, Alemania, Reino Unido, Italia y España formaron el consorcio "Eurofighter". Son también de interés las informaciones sobre el reparto de trabajo entre los grupos involucrados, así como las estimaciones de costes.

</ES-narr>

</top>

#### Beispiel 4: Erster finnischer EU-Kommissar

<top>

<num> C118 </num>

<FR-title> Le premier commissaire européen finlandais à l'UE </FR-title>

<FR-desc> Quel personnage d'origine finlandaise a été élu commissaire européen à l'Union Européenne? </FR-desc>

<FR-narr> Donner le nom du premier Finlandais nommé commissaire européen. Les documents pertinents peuvent également faire mention des tâches et missions qui incomberont au nouveau commissaire. </FR-narr>

</top>

<top>

<num> C118 </num>

<FI-title> Suomen ensimmäinen EU-komissaari </FI-title>

<FI-desc> Kuka nimettiin Suomen ensimmäiseksi komissaariksi Euroopan Unioniin? </FI-desc>

<FI-narr> Kerro Suomen ensimmäisen Euroopan Unionin komissaarin nimi. Relevantteja ovat myös dokumentit jotka kertovat uuden komissaarin toimeen liittyvistä toimintaalueista. </FI-narr>

</top>

## 2.4 Pooling-Methode

Die Schritte des Bewertungsverfahrens, das auf der Pooling-Methode von TREC basiert, sind wie folgt: Die teilnehmenden Systeme liefern ihre integrierten, geordneten Ergebnislisten pro Themenstellung. Diese Listen enthalten die Nummern derjenigen Dokumente für die jeweiligen Themen, die von den Systemen als relevant ermittelt wurden, und zwar in der Reihenfolge der vermuteten Relevanz (absteigend). Die jeweils ersten 60 Dokumente werden in den Pool für den entsprechenden Topic einbezogen. Alle Ergebnislisten, die

zu einer der 50 Themenstellungen der Hauptaufgabe bzw. zu einer der je 25 Themenstellungen der domänenspezifischen Aufgabe (GIRT) und der wissenschaftlichen Aufgabe (MARYLLIS) gehören, werden zusammengespielt und in eine Zufallsreihenfolge<sup>3</sup> gebracht; dadurch ist nicht mehr feststellbar, welches Dokument von welchem System stammt oder an welcher Stelle der vermuteten Relevanzreihenfolge es stand. Dann werden die Listen nach Sprachen aufgeteilt: alle Dokumentnummern aus den Korpora, die zu einer bestimmten Sprache gehören, werden zusammengeführt. Auf diese Weise wird pro Themenstellung eine große Zahl von Dokumenten zur jeweiligen Sprache zur Relevanzbewertung gesammelt. Die zu bewertende Maximalanzahl von Dokumenten könnte  $n \cdot 60 \cdot 50$ <sup>4</sup> sein, aber tatsächlich kommt nur etwas mehr als ein Viertel dieser Anzahl zusammen, weil nicht alle Gruppen die Maximalanzahl von Treffern pro Thema und Sprache beitragen und manche Treffer von mehreren Systemen abgeliefert werden - also identisch sind. Für die beiden wissenschafts- bzw. fachbezogenen Aufgabenstellungen kann maximal jeweils eine separate Anzahl von  $n \cdot 60 \cdot 25$  erwartet werden, aber auch hier wird de facto nur etwas mehr als ein Viertel dieser Anzahl erreicht.

## **2.5 Relevanz-Bewertung**

Diese gemeinsamen, geordneten Ergebnislisten pro Sprache werden von den Juroren der jeweiligen Sprachgruppe beurteilt. Die Beurteilung wird mit Hilfe der von NIST entwickelten Bewertungssoftware ASSESS festgehalten. Die Entscheidungen der Bewerter (als relevant oder nicht relevant in Hinsicht auf das jeweilige Thema) werden den sprachbezogenen Ergebnislisten jeder Themenstellung hinzugefügt. Die Bewerter wenden allgemeine Beurteilungsregeln an, die mit denjenigen von TREC vergleichbar sind; sie ziehen die Themendiskussionen der Sprachgruppen als Richtlinien für die Relevanzentscheidungen in Betracht und sie benutzen die Narratives der Themenstellungen als Entscheidungshilfen.

Trotzdem ist die binäre Entscheidung manchmal schwierig und häufig äußern die Juroren den Wunsch nach einer abgestuften Skala von Relevanz. Wie TREC hält aber auch CLEF aufgrund der besseren Auswertbarkeit an der binären Relevanzentscheidung fest.

---

<sup>3</sup> Dies ist die aufsteigende Reihenfolge nach den Dokumentnummern der jeweiligen Korpora.

<sup>4</sup> n (Anzahl von teilnehmenden Gruppen)  
60 (zutreffende relevante Dokumente pro Themenstellung)  
50 (Anzahl von Themenstellungen).

## 2.6 Aufgabenstellungen (Tasks)

CLEF zielt auf die Weiterentwicklung von multilingualen IR-Systemen ab. Auf dem Weg zum Testen mehrsprachiger IR-Systeme sollen jedoch auch Schritte zur Einbeziehung weiterer Sprachen und zum Sammeln von Erfahrungen bei der Ausrichtung von Tests möglich sein. Daher werden unterschiedliche Aufgabestellungen (Tasks) formuliert, denen sich die Teilnehmer stellen können<sup>5</sup>.

Die Hauptaufgabenstellung von CLEF ist das **mehrsprachige IR** (multilingual task):

Suche nach Dokumenten in allen Hauptsprachen, wobei eine dieser Sprachen als Anfragesprache dient, sowie Erstellung einer integrierten Liste aller Ergebnisse aus allen Dokumentsammlungen (d.h. aus allen Hauptsprachen).

Zusätzlich werden weitere Sprachen als Ausgangssprachen möglich gemacht, da entsprechende Übersetzungen der Themenstellungen von teilnehmenden Gruppen erstellt werden (Finnisch, Russisch, Schwedisch). Die Hauptsprachen bleiben dabei ebenfalls Zielsprachen.

Die **zweisprachige Aufgabenstellung** (bilingual task) sieht vor, in einer beliebigen Ausgangssprache (die nicht gleich der Zielsprache ist) nach Dokumenten in englischer oder holländischer Sprache zu suchen. Aus diesem Grund stellten die CLEF-Organisatoren auch Übersetzungen der Themenstellungen ins Holländische sowie weitere linguistische Ressourcen für das Holländische (Stoppwortliste, Stemmer, Holländisch-Englisch Lexikon) zur Verfügung.

Die **einsprachige Aufgabenstellung** (monolingual task) sieht vor, in deutscher, englischer, französischer, holländischer, italienischer und spanischer Sprache nach Dokumenten in einer der entsprechenden Dokumentsammlungen zu suchen. Englisch wird dabei ausgeschlossen, da Englisch schon in der Vergangenheit durch den Ad-hoc-Retrieval Task von TREC abgedeckt wurde und keine neue Herausforderung im Hinblick auf linguistische Probleme und Übersetzungsfragen darstellt. Die einsprachige Aufgabenstellung ist einerseits als Einstieg für neue CLEF-Teilnehmer gedacht, andererseits können auf diese Weise neue Sprachen für die multilinguale Aufgabenstellung eingeführt werden.

---

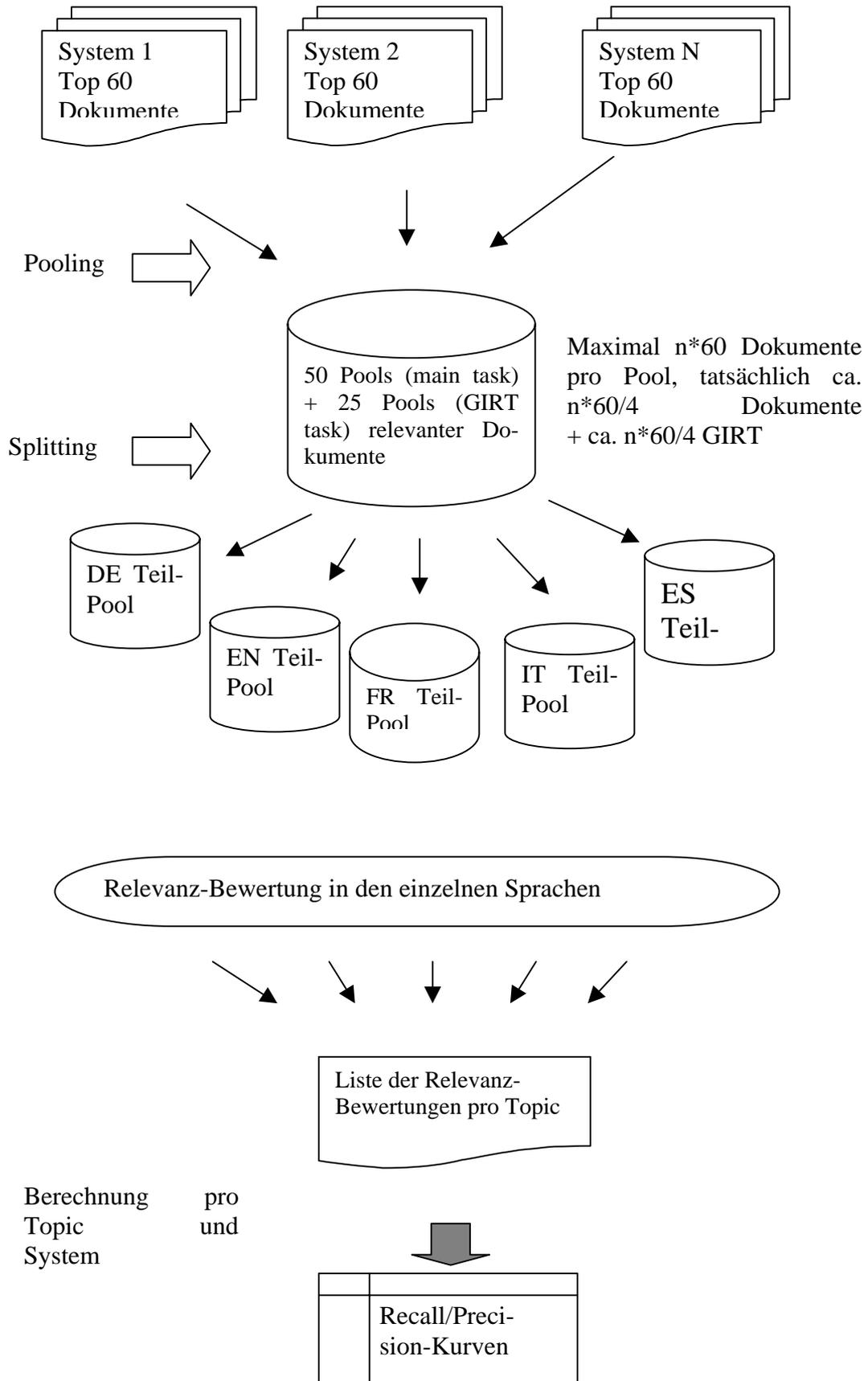
<sup>5</sup> <http://www.iei.pi.cnr.it/DELOS/CLEF/clef00.html>.

Die wissenschaftliche bzw. **fachbezogene Aufgabenstellung** (scientific and domain-specific task) erlaubt, nach (sozial)wissenschaftlichen Dokumenten in speziellen Dokumentensammlungen, nämlich GIRT (German Indexing and Retrieval Testdatabase) oder AMARYLLIS zu suchen. Damit reagiert CLEF auf den oft genannten Vorwurf, große Evaluierungen nur auf der Basis von Zeitungstexten führten nicht zu übertragbaren Ergebnissen. Die Dokumente der GIRT- und AMARYLLIS -Datenbanken enthalten auch intellektuell vergebene Schlagwörter jeweils aus einem (sozial)wissenschaftlichen Thesaurus, der ebenfalls zur Verfügung gestellt wird, und zwar auch in englischer bzw. bei GIRT zusätzlich noch in russischer Übersetzung. Außerdem werden dafür spezifische Themenstellungen auf Englisch und Deutsch bzw. Französisch (GIRT außerdem Russisch) bereitgestellt. AMARYLLIS und GIRT bieten eine ideale Plattform, um die Übertragbarkeit der Systeme auf spezifische wissenschaftliche Fachtexte zu testen (cf. Gey & Kluck 2001).

Zusätzlich wurde eine experimentelle Aufgabenstellung definiert: die **interaktive Aufgabenstellung** (interactive task). Ziel dieser Aufgabe ist es, die Evaluierung von interaktivem CLIR zu erforschen und Vergleichsmaßstäbe zu entwickeln, an denen weitere Forschungen gemessen werden können. In diesem Track wird also die Retrievaleffektivität in Kombination mit der Benutzungsoberfläche bewertet. Dabei geht es insbesondere um die Möglichkeiten, die Anfrage zu formulieren und zu verändern und die Ergebnisdokumente schnell bewerten zu können. Die Anfragen werden in diesem Fall von Testpersonen bearbeitet und nicht automatisch vom System oder von Experten erstellt (cf. Gonzalo & Oard 2002).

## **2.7 Bearbeitung der Themenstellungen durch die Teilnehmer**

Die von den Teilnehmern verwendeten Retrievalsysteme beginnen mit der Suche nach einer Themenstellung in einer Sprache (z.B. DE) und liefern Dokumente in allen Zielsprachen zurück. Für die Suchen in den Dokumentbeständen (in der Hauptaufgabenstellung des mehrsprachigen IR: DE, EN, ES, FR, IT) verwenden sie systemspezifische Strategien, um die Übersetzung oder Transformation der Fragen in andere Sprachen zu lösen. Am Ende des Rückgewinnungsprozesses müssen sie einen integrierten und geordneten Ergebnissatz der Dokumente liefern, von denen angenommen wird, dass sie für die Themenstellung relevant sind. Der Prozess der Integration der Ergebnisse aus verschiedenen Datenbeständen (merging) ist neben der Lösung der Übersetzungsprobleme eine zweite nicht unwichtige Herausforderung.



### **3 Die CLEF-Kampagne 2002: Überblick**

Der Überblick wird die folgenden Abschnitte behandeln und aktuell vom CLEF 2002 Workshop im September 2002 in Rom berichten:

- Teilnehmer und verwendete Sprachen
- verwendete Verfahren zum CLIR
- verwendete Ressourcen für die Bewältigung der Multilingualität
- interaktives CLIR
- fachbezogenes CLIR: Amaryllis und GIRT

### **4 Forschungsfragen des CLIR und der Evaluation**

Wesentliche Forschungsfragen im Bereich CLIR sind unter anderem folgende:

- die Kombination geeigneter Methoden zur Behandlung der Multilingualität

Wie die Ergebnisse der Ad-Hoc-Tasks von TREC zeigten, lassen sich mit der Kombination verschiedener Varianten eines Retrievalsystems oder mit der Kombination verschiedener Retrievalsysteme durchaus Steigerungen der Leistung erzielen. Dazu liegt auch der Vorschlag zur Kombination im Rahmen des MIMOR-Modells vor. Jedoch sind diese Kombinationen nicht beliebig, sondern von den spezifischen Eigenschaften der Komponenten oder Systeme abhängig, die kombiniert werden. Hier ist immer noch ein weites Feld empirischer und systematischer Forschung.

- die Integration der Ergebnisse in eine Ergebnisliste (Merging)

Das Zusammenführen der gerankten Ergebnisse aus verschiedenen Datenbeständen (in verschiedenen Sprachen) ist kein triviales Problem. Die Spanne der vorhandenen Ansätze reicht von ganz mechanischen (und daher auch theoretisch völlig unbefriedigenden, aber in der Praxis durchaus erfolgversprechenden) Verfahren zu ausgeklügelten Überlegungen hinsichtlich der Normierung und Integration unterschiedlicher RSV der Einzelergebnisse in das endgültige Ranking.

- die Präsentation der Ergebnisse für den Benutzer (einschließlich multilingualer Hilfsmittel)

Hier lassen sich verschiedene Verfahren der Ergebnispräsentation wie automatische Zusammenfassungen und deren Übersetzung oder die Wiedergabe

wichtiger Sätze oder Phrasen in der Anfragesprache des Benutzers denken. Aber auch die Visualisierung von Ergebnismengen fiele in diese Kategorie.

- die Interaktion von System und Benutzer

Die Interaktivität von Systemen bietet einen weiteren Gesichtspunkt der nach wie vor intensiver Forschung bedarf, hierbei stellt die Multilingualität eine zusätzliche Problemebene dar.

- die Weiterentwicklung der Evaluationsmethodik

Die Methodologie der Evaluation ist sowohl in Bezug auf die Poolingmethode als auch im Hinblick auf neue Aufgabenstellungen des CLIR (Websearch, Question & Answering, CLIR im Multimedia-Kontext usw.) weiterzuentwickeln und weiterhin kritisch zu durchleuchten.

- die vergleichende Analyse der vorgelegten Resultate aus den vergangenen Tests

Nunmehr liegen die Ergebnisse einer Reihe von Evaluationskampagnen vor, die eine Vielzahl von Daten lieferten, die der weiteren Analyse harren. Insbesondere die vergleichende Analyse ähnlicher Verfahren hinsichtlich ihrer Spezifika und der Wirkungen bestimmter Ausprägung der Systeme.

## **5 Ausblick auf die weitere Entwicklung von CLEF**

Gegenwärtig wird die weitere Entwicklung des CLEF nach Ablauf der bestehenden Projektförderung durch die EU-Kommission diskutiert. Für die Planungen des 6. Forschungsrahmenprogramms der EU wurde eine Interessenbekundung für ein Exzellenznetzwerk abgegeben, an der sich eine Vielzahl an Instituten und Firmen aus ganz Europa beteiligt hat. Generell geht es um die Einbeziehung weiterer Sprachen in die Hauptaufgabenstellung des multilingualen Retrieval, zunächst werden wohl Niederländisch, Schwedisch und Finnisch hinzukommen. Wichtig ist auch die Erweiterung um Sprachen aus unterschiedlichen Sprachfamilien wie z.B. Russisch (slawische Sprache mit kyrillischen Zeichen), eine weitere slawische Sprache mit lateinischen Zeichen (aber Ligaturen), eine weniger gesprochene bzw. regionale Sprache wie Baskisch oder Rätoromanisch.

## **6 Literaturangaben**

- Braschler, M., 2002: CLEF 2002 - Overview of Results. In: C. Peters (Hrsg.) Workshop of the Cross-Language Evaluation Forum, CLEF 2002, Rome, Italy, September 2002, Sophia-Antipolis: ERCIM (erscheint)
- Braschler, M., 2002: CLEF 2001 - Overview of Results. In M. Braschler, J. Gonzalo, M. Kluck, C. Peters (Hrsg.), Evaluation of Cross-Language Information Retrieval Systems. Second Workshop of the Cross-Language Evaluation Forum, CLEF 2001. Darmstadt, Germany, September 2001. Revised papers, Berlin et. al: Springer (LNCS 2406)
- Braschler, M., 2001: CLEF 2000 - Overview of Results. In: C. Peters (Hrsg.), Cross Language Information Retrieval and Evaluation, Workshop of the Cross-Language Evaluation Forum, CLEF 2000, Lisbon, Portugal, September 2000, Revised Papers. Berlin et al.: Springer (LNCS 2069), 89-101
- Harman, D., M. Braschler, M. Hess, M. Kluck, C. Peters, P. Schäuble, and P. Sheridan, 2001. CLIR Evaluation at TREC. In: C. Peters (Hrsg.), Cross Language Information Retrieval and Evaluation, Workshop of the Cross-Language Evaluation Forum, CLEF 2000, Lisbon, Portugal, September 2000, Revised Papers. Berlin et al.: Springer (LNCS 2069), 7-23
- Kluck, M., T. Mandl and C. Womser-Hacker 2002. Cross-Language Evaluation Forum (CLEF) – Europäische Initiative zur Bewertung sprachübergreifender Retrievalverfahren. In: Information Wissenschaft & Praxis 53:82-89
- Kluck, M., C. Womser-Hacker 2002: Inside the Evaluation Process of the Cross-Language Evaluation Forum (CLEF): Issues of Multilingual Topic Creation and Multilingual Relevance Assessment. In: M. G. Rodríguez, C. P. S. Araujo (Hrsg.): Proceedings of the Third International Conference on Language Resources and Evaluation, LREC 2002, Las Palmas de Gran Canaria 29-31 May 2002, Paris: ELRA, 573-576 preprint: <http://www.educat.hu-berlin.de/~kluck/lrec-paper-final.doc> oder <http://www.educat.hu-berlin.de/~kluck/lrec-paper-final.pdf>
- Voorhees, E., 2001. Philosophy of IR Evaluation. In: C. Peters (Hrsg.): Results of the CLEF 2001 Cross-Language System Evaluation Campaign, Working Notes for the CLEF 2001 Workshop, 3 September, Darmstadt, Germany, Sophia-Antipolis: ERCIM, 257-260
- Voorhees, E., 2000. Variations in Relevance Judgements and the Measurement of Retrieval Effectiveness. Information Processing & Management 36:679-716
- Voorhees, E. and D. Harman, 2001. Overview of the Ninth Text Retrieval Conference (TREC-9). In: E. Voorhees, D. Harman (Hrsg.) The Ninth Text REtrieval Conference (TREC 9). Gaithersburg: NIST, 1-14, [http://trec.nist.gov/pubs/trec9/t9\\_proceedings.html](http://trec.nist.gov/pubs/trec9/t9_proceedings.html)
- Womser-Hacker, Ch. (1996), Das MIMOR-Modell. Mehrfachindexierung zur dynamischen Methoden-Objekt-Relationierung im Information Retrieval. Habilitationsschrift, Universität Regensburg.