Evaluierung von freeWAIS-sf

Evaluierung durch

- Kai Grossjohann (Uni Dortmund)
- Pilotanwender:
  - Uwe Berger (Uni Stuttgart)
  - Patrick Zeimetz (IBFI Dagstuhl)
  - M. Eimermacher (UBIS GmbH)
  - Claus Gremminger (FH Karlsruhe)
  - P. Schmitt (Uni Karlsruhe)
  - Rainer Lienhart (Uni Mannheim)

Technische Eigenschaften

- Anzahl der Nutzer nur durch Leistung des Rechners beschränkt
- Dokumentgröße nur indirekt beschränkt, da Terme, die häufiger als 20.000 mal vorkommen, als Stoppwörter betrachtet werden.
- Streßtest (mit News-Spool) ergab Probleme der aktuellen Version von freeWAIS-sf mit Datenmengen über 100 MB beim Suchen (vermutlich Programmierfehler, keine prinzipiellen Probleme)
- Speicherplatzbedarf (Plattenplatz):
  - Software (freeWAIS-sf, SFgate): < 10 MB
  - MeDoc-Suite: Index ca. so groß wie Daten
  - TIME-Suite: Index ca. 2,5 mal so groß wie Daten
  - News: Index ca. 1,2 mal so groß wie Daten

- Geschwindigkeit beim Einfügen und Indexieren:
  - MeDoc-Suite: 100 Dokumente, 8,5 MB, 120s
  - TIME-Suite: 423 Dokumente, 1,5 MB, 90s
  - News: 160.000 Dokumente, 300 MB, 7 Stunden

  Einfügen eines einzelnen Dokuments im Sekundenbereich.
- Geschwindigkeit beim Retrieval:
  Bei Datenbank auf lokaler Platte deutlich unter einer Sekunde, bei Zugriff über Netz unter zwei Sekunden.
- Verhalten im Internet:
  WWW-Interface (Perl-Skript) hat lange Startzeit, dadurch Antwortzeit typisch ca. 20s.

Dokumentenzugriff

- Retrievalqualität
  - Precision: nach 5 Dokumenten 0.26, nach 10 Dokumenten 0.18
  - Recall: nach 5 Dokumenten 0.44, nach 10 Dokumenten 0.52
- Dokumentenformate und -darstellung
  - Indexierung
    * nur ASCII-Text (auch ISO-8859-1)
    * Text kann Feldkennungen enthalten
    * andere Formate müssen über Filter in diese Darstellung konvertiert werden
  - Darstellung der Dokumente (SFgate, WWW):
    MIME-Mechanismus

Administrative Eigenschaften

- Installation
  - Installation der Software durch Skript gesteuert, das einige Informationen erfragt (Directories, Varianten).
  - keine nennenswerten Probleme aufgetreten
- Einrichten einer Datenbank
  - Erstellung der Format-Beschreibung für die Indexerstellung sehr flexibel aber auch aufwendig (insbesondere die Einarbeitung).
  - Tool zur Prüfung der Syntax der Format-Beschreibung
  - Tool soll in Zukunft auch Erstellung der Format-Beschreibung unterstützen

- Administration im laufenden Betrieb
  - Tool `makedb` zum Hinzufügen von Dokumenten
  - Datenbank bleibt während der Aktualisierung benutzbar
  - `makedb` für regelmäßige Änderungen (z.B. jede Nacht), weniger für sofort wirksame Änderungen.

Beurteilung des Herstellers

- freeWAIS-sf an Uni Dortmund im Rahmen einer Diplomarbeit aus freeWAIS entwickelt
- Wartung inzwischen durch Firma WSC Technologies
- für MeDoc-Projekt Wartung durch Mitarbeiter der Uni Dortmund
- freeWAIS-sf auch weiterhin kostenfrei von WSCT verfügbar

Brauchbarkeit für MeDoc

- Schnittstellen:
  - direkter Zugriff auf lokale Indexdaten:
    schnell, Zugriff auf Metadaten
  - WAIS-Protokoll:
    kein Zugriff auf Metadaten aus der WAIS-Datenbank (durch Erweiterung des WAIS-Protokolls wäre auch dies möglich)
- Benutzerverwaltung und Accounting
  - keine Nutzerverwaltung
  - Log-Files mit Protokoll der Zugriffe (IP-Adresse)
- Annotationen: werden nicht unterstützt

- Sichere Kommunikation:
  bei Verwendung von SFgate erfolgt Kommunikation über HTTP, es kann also eine sichere Variante von HTTP eingesetzt werden

Zusammenfassung

- gute Retrievalqualität
- ansprechende WWW-Oberfläche (SFgate)
- frei verfügbar
- Schwächen:
  - Nutzerverwaltung und Accounting
  - Updates im laufenden Betrieb