PEFNA

The Private Filtering News Agent

Dpt of Computer and Systems Sciences Stockholm University

Frederik Kilander
Eva Fahraeus
Jacob Palme

Was ist PEFNA?

- Prototyp eines Hilfsmittels zum Lesen von Usenet News. (Stand: Feb. '97)
- unterstützt den Benutzer durch eine Vorklassifizierung der Nachrichten.
  - Nachrichten werden in (benutzerdefinierte) Kategorien sortiert.
  - Nachrichten einer Kategorie werden nach ihrem Ähnlichkeitsmaß sortiert
    - ähnlichste Nachricht steht am Anfang der Liste.
- PEFNA besteht aus zwei Komponenten.
  - PFCLIENT (Newsreader)
    - X- basiertes Tcl/Tk- Programm
  - PFAGENT (in C implementiert)
- ist personengebunden. Informationen werden unter *$HOME/.xnewsrc* abgelegt

Systembeschreibung

Benutzerdaten Ablage

- Unterverzeichnis in *$HOME*
  - Artikelklassifikationsdatenbank (ASCII)
    - Message- ID, NewsGroup, Seq.-Nr (NNTP), Flags, Datum
  - Beispielverzeichnisbaum
    - “gute Beispiele”
    - Wortliste des Agenten
  - Protokolldatei

Systembeschreibung

Der Agent

- vergleicht *.xnewsrc* mit NNTP- Server
- legt neue Daten in der Datenbank ab
- gelesene & klassifizierte Artikel “verfallen” nach 1 Woche
- Reanalyse nach Verändern der Kriterien
    - d.h. durch neue “gute Beispiele”
- Analyse durch zwei Algorithmen
    - Zitatanalyse
    - Dokumentenähnlichkeit
- modularer Aufbau

Algorithmen

Zitat- Analyse

- Textlänge
- Länge der Zitate vs Länge der Kommentare
- Anzahl der Wechsel zwischen Zitat und Text
- spezielle Terminologien
    - ***"In article** message-ID auther **wrote:"***

Algorithmen

Kategorie- Ähnlichkeit nach Salton (Cosinus Maß)

- Vergleich: jeder Artikel gegen jedes Beispiel

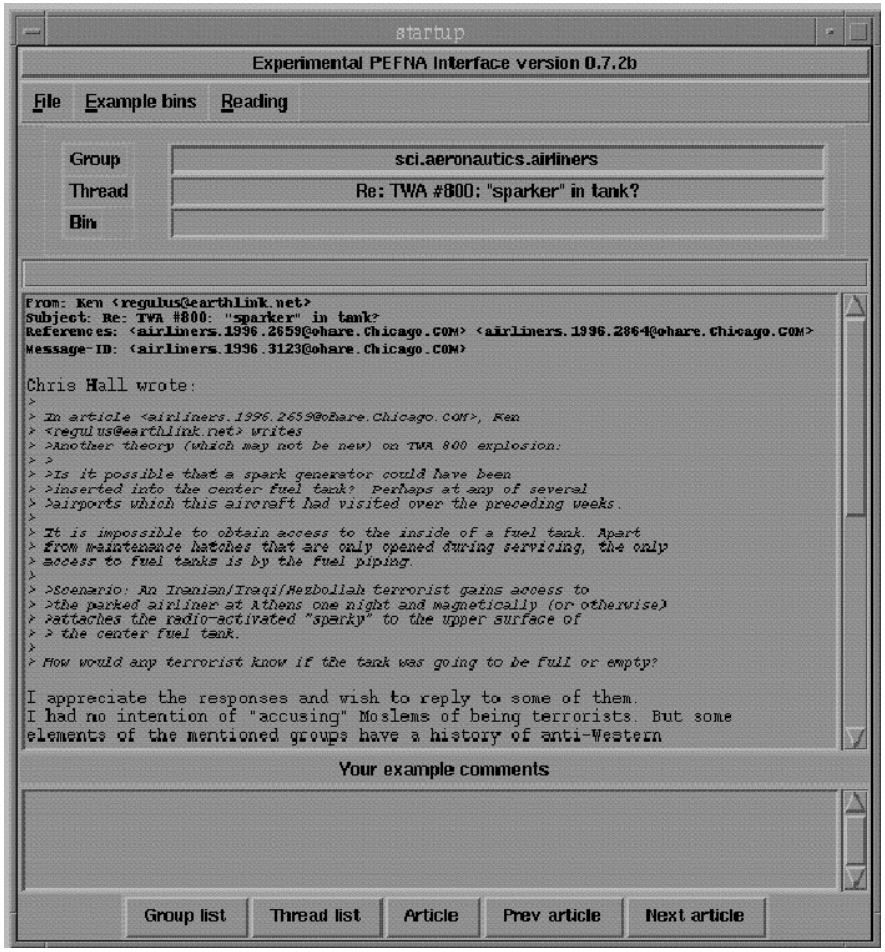
Zur Erinnerung:

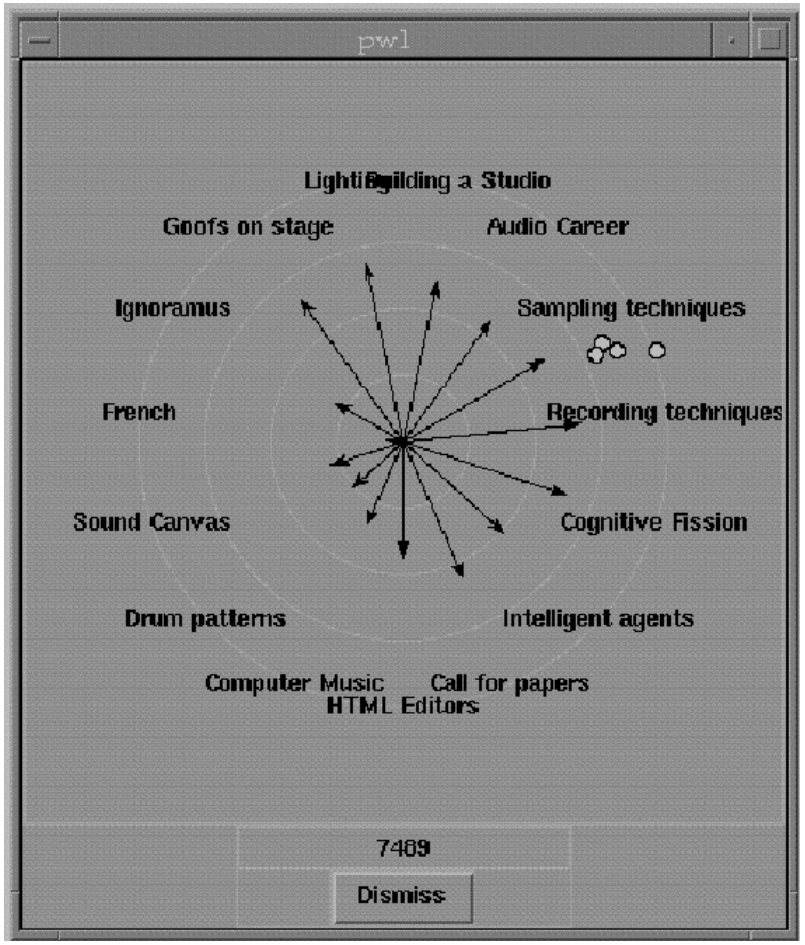
$$sim_{a,c} = \frac{\sum_t w_{a,t} \cdot w_{c,t}}{\sqrt{\sum_t w^2_{a,t}} \cdot \sqrt{\sum_t w^2_{c,t}}}$$

$w_{a,t}$:= Gewicht eines Wortes im Artikel (Häufigkeitsmaß)
$w_{c,t}$:= Gewicht eines Wortes in der Kategorie

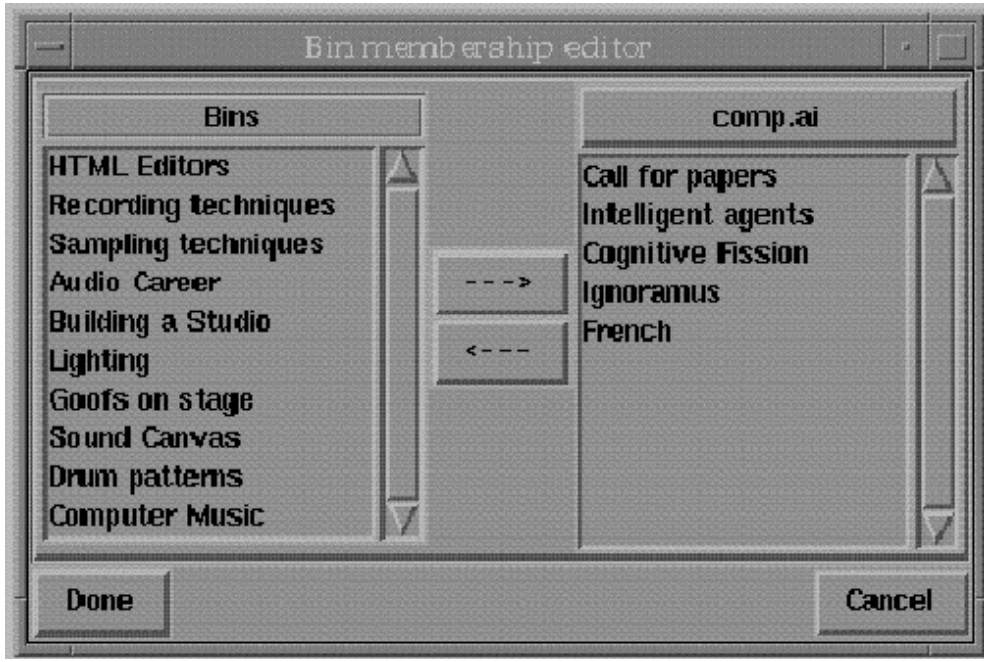
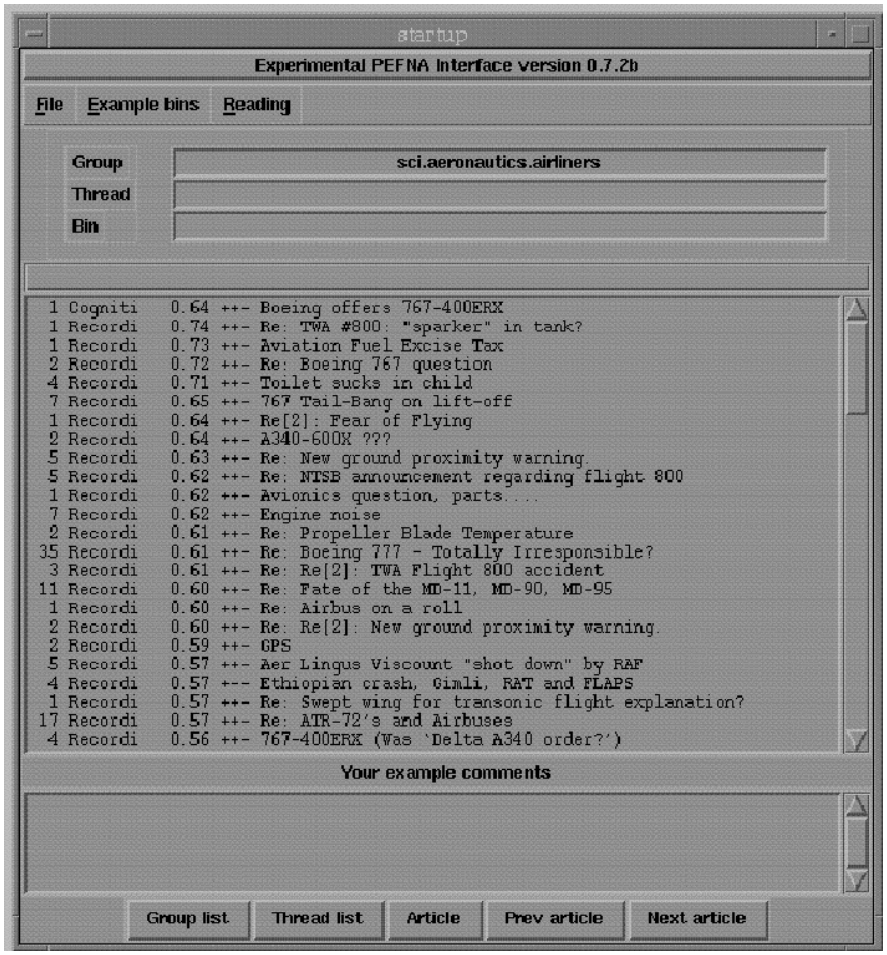
- Lange Laufzeiten

Screenshots





Screenshots



Screenshots