Referat zum Thema

Agenten zur Unterstützung beim Browsen auf dem WWW

von

Hendrik Schreiber

Letizia - An Agent that assists Web Browsing

- entwickelt von Henry Lieberman (MIT)
- autonomer Interface-Agent
- autonom, da Letizia absolut selbständig arbeitet, d.h. ohne jegliche Bestätigung seiner Aktionen vom Nutzer.
- Interface, da Letizia dem Nutzer aktiv bei der Bedienung eines Interfaces hilft, nämlich dem Web (anderes Beispiel: Online-Hilfe).

Was macht Letizia?

- Letizia versucht für seinen Nutzer interessante Dokumente zu finden, indem es - ausgehend von der Position des Nutzers - den angebotenen Links folgt. Findet es relevante Seiten, schlägt es diese dem Nutzer unaufdringlich vor.
- Die Bewertung der Seiten wird vom „Surf“-Verhalten des Nutzers abgeleitet.

Wie folgt Letizia den Links?

- Breitensuche im Gegensatz zu der beim „normalen“ Browsen favorisierten Tiefensuche.
- Letizia stoppt bei Seiten die Formulare enthalten.
- Es gibt ein Ressourcen-Limit: maximale Anzahl an besuchten, nicht-lokalen Seiten pro Minute.

Wie ordnet Letizia Dokumente ein?

Interessante Dokumente sind jene,

- die in einer Bookmark-Liste gespeichert wurden.
- die viele Links enthalten, denen der Nutzer gefolgt ist.
- die sich der Nutzer relativ zu ihrer Länge lange anguckt.

Uninteressante Dokumente sind jene,

- über die der Nutzer „hinwegsurft“.

Worauf basieren Letizias Empfehlungen?

Letizia benutzt TFIDF.

Lira

- wurde entwickelt von Marko Balabanovic und Yoah Shoham (Stanford)
- soll den Nutzer über interessante Seiten auf den Laufenden halten.

Dabei wird folgender Algorithmus benutzt:

- Durchsuche das Netz in einer begrenzten Zeit und benutze dazu eine geeignete Heuristik.
- Bewerte die Seiten mit einer Heuristik, welche die gleiche sein kann, wie unter Punkt 1 und präsentiere dem Nutzer die p besten Seiten.
- Erhalte eine Bewertung der Seiten von dem Nutzer.
- Bringe die Such- und Bewertungsheuristiken entsprechend den Bewertungen auf den neusten Stand.

Wie sucht Lira die Seiten?

Best-First, dabei wird die Güte einer Seite durch die Bewertungs-Heuristik beurteilt.

Wie werden die Seiten bewertet?

- Lira benutzt dokumentenlängen-normiertes TFIDF (vgl. Salton & Buckley, 1987) basierend auf einem 27.000 Wörter umfassenden Dictionary sowie Stemming.
- Die Dokumentenvektoren enthalten jeweils nur die zehn am höchsten gewichteten Wörter (mehr Wörter würden das Ergebnis auch nicht wesentlich verbessern, vgl. Salton & Buckley, 1987).
- Vor der Bewertung werden alle HTML-Tags entfernt.

Wie lernt Lira?

Relevance Feedback (nach Rocchio 1971).

$$\vec{M} \leftarrow \vec{M} + \sum_{i=1}^{p} e_i \vec{V}_i$$

dabei gelte:

e_i: ganze Zahl zwischen -5 und +5

V_i: Dokumentenvektor

M: Vektor der Präferenz-Klasse

Was nach dem Testen übrig blieb...

- In einem 24 Tage dauernden Test schnitten die von Lira empfohlenen Dokumente nach kurzer Zeit besser ab als zufällig ausgesuchte Seiten.
- Insgesamt wurden die von Lira vorgeschlagenen Seiten in fünfzig Prozent der Fälle höher bewertet als die von Menschen empfohlenen „Cool-Sites“.
- Da sich das System auf ein Thema einschießt, verzerren die Nutzer ihre Bewertungen absichtlich.