

Content management: T-map

Keywords

Thesaurus, ontologieën, classificatie, kennistechnologie, indexering

Probleemstelling

Het indexeren van een documentcollectie vergroot de toegankelijkheid. T-map vult full-text indexering aan op twee wijzen: 1. Via terminologie-extractie krijgt de gebruiker een idee waar de documentenverzameling over gaat. 2. De documenten met vergelijkbare terminologie worden geclusterd zodat de gebruiker veel efficiënter overzicht krijgt. Op deze wijze kan T-map een aanzienlijke versnelling leveren bij bijvoorbeeld het bouwen van thesauri en ontologieën.

Technisch probleem

Hoe moet je documenten clusteren als je van tevoren niet weet hoeveel verschillende klassen er zijn, of wat de inhoud van deze klassen is? Hoe extraheer je de terminologie zonder redundantie waarbij je rekening houdt met ambiguïteit, synonymie, etc.

Soort oplossing

Een zelfstandig systeem of halffabriek. Het product zal aansluiten bij een bestaand systeem. Momenteel is het systeem beschikbaar als demonstrator. De technologie kan ook worden toegepast op transcripties van audiomateriaal die verkregen zijn met behulp van automatische spraakherkenning. (Zie onder Audio Watcher).

Vergelijking met bestaande oplossingen

De beschikbare systemen verrichten alleen een terminologie-extractie. Veelal wordt onvoldoende gebruik gemaakt van natuurlijke taalanalyse waardoor er veel redundantie ontstaat. Daarnaast worden de documenten niet geclusterd.

Potentiële afnemers

- Alle bedrijven die met een thesaurus werken.
- Inlichtingendiensten.
- Marketingbureaus.
- Toeleveranciers van helpdesks.

Contact

TNO

TPD

kraaij@tpd.tno.nl