

Leren: plaatjes in publicaties

Keywords

Zoeken, beelden, informatie systemen, content management, zoeken, leren

Probleemstelling

Uitgevers zijn in de afgelopen jaren massaal overgegaan op internet gebaseerde digitale bibliotheken. Deze sites zijn echter een directe afspiegeling van het bestaande papieren tijdschrift. In nieuwe vormen van publiceren zullen vooral de foto's en figuren een grote rol moeten gaan spelen. Echter om te zoeken in grote verzamelingen van beelden is indexering op genres van plaatjes van essentieel belang.

Technisch probleem

De beschrijvingen van beelden zijn impliciet in het beeld en de beschrijving in de vorm van een caption. Sommige informatie is exclusief in het beeld, andere informatie in het bijschrift, vaak een combinatie van beide.

Soort oplossing

Er is een prototype gebouwd voor het biomedische domein. In het ontwikkelde systeem wordt op basis van beeldkenmerken elk beeld automatisch in een bepaalde beeldcategorie geplaatst bijv. grafiek, micrograph, electronen-microscop beeld etc. Het bijschrift wordt geanalyseerd met een techniek genaamd Latent Semantic Indexing. Dit is een analyse techniek die veel verder gaat dan het kijken naar de keyword. Elk bijschrift wordt automatisch gekoppeld aan relevante concepten, ook als het concept in die bewoording niet in het bijschrift voorkomt.

Vergelijking met bestaande oplossingen

Soortgelijke benaderingen zijn nog niet gedaan voor biomedische beelden. Wel zijn er combinaties van bijschriften en krantefoto's bekend, maar deze zijn niet overdraagbaar naar het biomedische domein.

Potentiële afnemers

Uitgevers (Elsevier), Onderzoekslaboratoria (Unilever).

Contact

Universiteit van Amsterdam

Informatica – ISIS

<http://www.science.uva.nl/research/isis/>

