

Antwoordpresentatie in een QA systeem:

Mag het ietsje meer zijn?

Mariët Theune, Emiel Kraemer**, Wauter Bosma*,
Erwin Marsi** en Charlotte van Hooijdonk***

** Universiteit Twente, Enschede*

*** Universiteit van Tilburg, Tilburg*

Lang heeft bij QA de nadruk gelegen op het beantwoorden van 'weetjesvragen' met een kort en simpel antwoord. ("Wat is de hoofdstad van Malawi?" "Lilongwe.") Tegenwoordig wordt echter van QA systemen verwacht dat zij ook meer diepgaande vragen aankunnen, wat onder meer als gevolg heeft dat er niet meer met een simpel feit als antwoord kan worden volstaan. Vragen als "Wat is het verschil tussen een vlokcentest en een vruchtwaterpunctie?" kunnen niet met een enkele frase beantwoord worden. Bij weer andere vragen ("Hoe kan ik mijn werkplek ergonomisch inrichten?") is een afbeelding vaak informatiever dan een puur tekstueel antwoord. Tenslotte is het gewenst dat de invoer- en uitvoermogelijkheden van een QA systeem met elkaar in balans zijn: als je in gesproken taal een vraag aan het systeem kan stellen, verwacht je ook dat het systeem terugpraat.

Het IMOGEN project richt zich daarom op een geavanceerde presentatie van antwoorden in een QA systeem, met als aandachtspunten 1) het formuleren van uitgebreide antwoorden, die informatiever zijn dan een simpele zin of frase, en 2) het genereren van multimediale antwoordpresentaties, waarbij tekstuitvoer gecombineerd wordt met afbeeldingen en spraak. Het doel is om te komen tot een antwoord dat beter aansluit bij de informatiebehoeften en voorkeuren van de vragensteller dan een eenvoudig 'standaard' antwoord. Om na te kunnen gaan of dit doel werkelijk bereikt wordt, is ook evaluatie een belangrijk onderdeel van het project.

Meer informatieve antwoorden

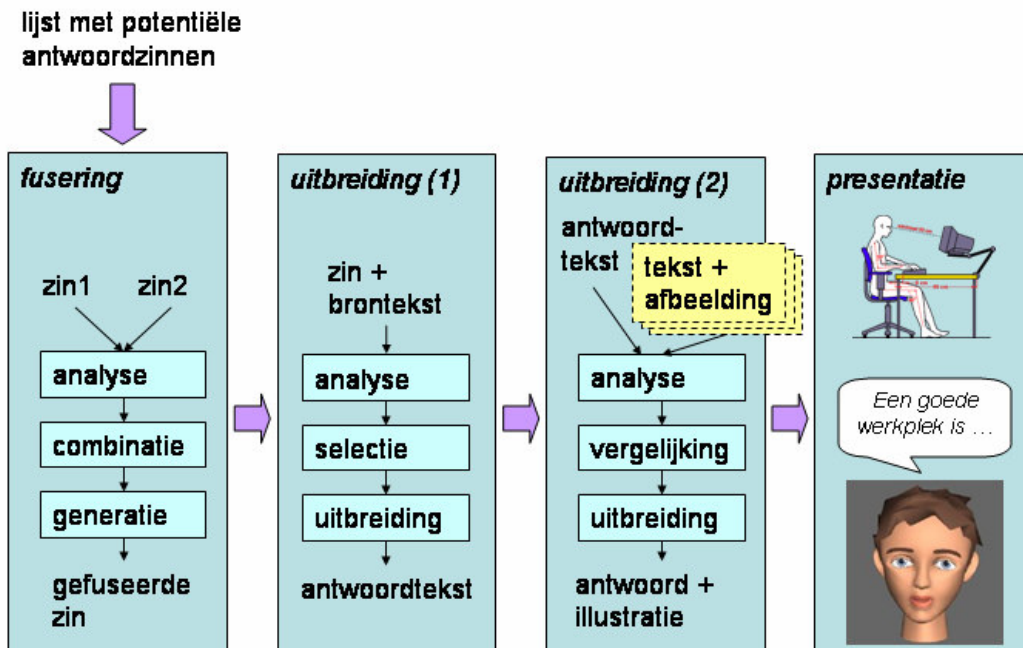
IMOGEN dient gebruikt te worden in combinatie met een QA-systeem, dat naar aanleiding van een gebruikersvraag een lijstje aanlevert met zinnen uit diverse documenten, die allemaal een potentieel antwoord bevatten op de vraag van de gebruiker. Binnen IMOGEN worden technieken ontwikkeld om één of meer van deze antwoorden op een optimale manier aan de gebruiker te presenteren. De eenvoudigste manier om dit te doen is door één van de zinnen uit het lijstje te selecteren en deze aan de gebruiker te tonen. In veel gevallen zal dit echter geen optimaal informatief antwoord opleveren. Zo kan een vraag als "Is RSI te genezen?" in principe afdoende beantwoord worden met de zin "RSI is te genezen, mits je er vroegtijdig bij bent." Echter, de vragensteller zou waarschijnlijk meer tevreden zijn met een uitgebreider antwoord, waarin ook vermeld wordt *hoe* RSI dan wel genezen kan worden. Daarom worden binnen IMOGEN verschillende methoden onderzocht om antwoorden te verrijken door informatie uit verschillende potentiële antwoorden met elkaar te combineren (zinfusering), of door gerelateerde zinnen uit het brondocument aan een antwoordzin toe te voegen (antwoordextensie). Beide benaderingen worden hieronder besproken.

Zinfusering

De potentiële antwoordzinnen die worden opgeleverd door een QA module zullen in veel gevallen een grote onderlinge overlap vertonen (als de QA zijn werk goed heeft gedaan). Een vraag als, "Wat is de invloed van de werkplek op RSI?" zou onder andere de volgende antwoordzinnen kunnen opleveren:

- Een ergonomisch ingerichte werkplek met goed afgesteld meubilair is van belang om gezond te werken.
- Een goede werkplek is een directe voorwaarde voor een verantwoorde zithouding.

Deze antwoordzinnen kunnen gefuseerd worden in drie stappen: analyse, combinatie en generatie (zie figuur 1).



In de eerste stap wordt de structuur van de zinnen geanalyseerd. Door de resultaten van deze analyses met elkaar te vergelijken, is het mogelijk om die zinsdelen te detecteren die in betekenis aan elkaar gerelateerd zijn (Marsi en Krahmer, 2005ab). Daarbij kan een onderscheid gemaakt worden tussen verschillende betekenisrelaties: zinsdelen uit de beide zinnen kunnen identiek zijn aan elkaar ('werkplek'), synoniem zijn ('gezonde', 'verantwoorde'), specifiek of juist algemener zijn ('een ergonomisch ingerichte werkplek met goed afgesteld meubilair', 'een goede werkplek') of elkaar deels overlappen.

Op basis van deze informatie kunnen zinnen met elkaar gecombineerd worden en kan een nieuw antwoord gegenereerd worden (stappen twee en drie). Wanneer algemene zinsdelen worden vervangen door hun meer specifieke tegenhangers, zal dit leiden tot een nieuw antwoord met een grotere informatiedichtheid dan elk van de oorspronkelijke antwoorzinnen:

- Een ergonomisch ingerichte werkplek met goed afgesteld meubilair is een directe voorwaarde voor een verantwoorde zithouding.

In sommige gevallen kan het juist wenselijk zijn om een algemene samenvatting van de antwoorden te genereren door alleen generaliserende zinsdelen te combineren:

- Een goede werkplek is van belang om gezond te werken.

Dit levert een minder gedetailleerd, maar mogelijk wel betrouwbaarder antwoord op doordat alleen die informatie die alle potentiële antwoorden gemeenschappelijk hebben, er in is opgenomen.

Antwoordextensie

Een volgende manier om tot een rijker antwoord te komen, is om extra zinnen uit het brondocument aan de antwoordzin toe te voegen. Onderzoek heeft uitgewezen dat gebruikers van QA systemen de voorkeur geven aan een antwoord dat uit enkele zinnen (alinea) bestaat, boven slechts een enkele zin of frase. Behalve dat een langer antwoord meer informatie bevat, maakt de aanwezigheid van meer context het de vragensteller ook makkelijker om te beoordelen of het gegeven antwoord wel echt aansluit bij de vraag. Aangezien de kwaliteit van QA nog lang niet perfect is, is dit laatste geen overbodige luxe. Stel dat de gebruiker vraagt, "Wat is een goede werkplek ter voorkoming van RSI?" en het systeem de volgende antwoordzin vindt:

- Een goede werkplek is voor iedereen anders.

Dit antwoord is, hoewel juist, op zichzelf weinig informatief. Door ook de context van de zin te presenteren wordt de kans groter dat de informatiebehoefte van de gebruiker bevredigd wordt:

- Een goede werkplek is voor iedereen anders. Om RSI te voorkomen, dient uw werkplek op de maten van uw lichaam afgestemd te zijn. Stel bijvoorbeeld de armlenningen zo af, dat onder- en bovenarm een hoek van 90 graden vormen.

Bovendien ziet de gebruiker nu dat het antwoord wel degelijk over RSI gaat, en niet over heel andere zaken, zoals in het volgende geval:

- Een goede werkplek is voor iedereen anders. Het realiseren van uw ideale werkplek is afhankelijk van uw individuele wensen en eisen. Daarom biedt de Kantoormeubelgigant u advies op maat.

De eenvoudigste manier om een antwoordzin uit te breiden, is door het selecteren van een aantal zinnen uit de brontekst die grenzen aan de antwoordzin. Dit zijn echter niet altijd de voor het antwoord meest relevante zinnen. Daarom wordt binnen IMOGEN gekeken naar een meer intelligente manier om zinnen uit de brontekst te selecteren, op basis van een analyse van de hiërarchische structuur van de brontekst (Bosma 2005a). Bij deze analyse wordt aan zinnen die in deze structuur een sleutelpositie innemen ten opzichte van de antwoordzin, een hogere relevantiewaarde toegekend dan aan andere zinnen. Uiteindelijk wordt, afhankelijk van de gewenste lengte van het antwoord, een aantal zinnen met een hoge relevantiewaarde geselecteerd (dit hoeven niet noodzakelijkerwijs aangrenzende zinnen te zijn). Deze worden bij de oorspronkelijke antwoordzin gevoegd. Zie uitbreiding (1) in figuur 1 voor een grafische weergave van dit proces.

Uit gebruikersexperimenten is gebleken dat deze uitbreidingsmethode leidt tot antwoorden die nuttiger zijn en minder irrelevante informatie bevatten dan antwoorden die zijn ontstaan door simpelweg de omringende zinnen aan een antwoordzin toe te voegen.

Visuele informatie

Het spreekwoord luidt dat een afbeelding meer zegt dan duizend woorden. Bij vragen zoals "Wat is een goede werkplek?" lijkt een afbeelding van ergonomisch afgesteld meubilair inderdaad efficiënter te zijn dan een uitgebreide beschrijving. Het is dus nuttig om multimediale antwoordpresentaties te kunnen genereren, die naast tekst ook visuele informatie kunnen bevatten. Daarvoor is het wel nodig om te kunnen bepalen welke modaliteit wanneer de voorkeur heeft. Welke modaliteit is het meest efficiënt en effectief in het overbrengen van bepaalde informatie, en – ook niet onbelangrijk – aan welke modaliteit geven gebruikers subjectief de voorkeur?

Om een antwoord te krijgen op dergelijke vragen zijn binnen IMOGEN experimenten gedaan waarin informatie over RSI-oefeningen werd aangeboden in drie verschillende modaliteiten: tekst, statische afbeeldingen en filmpjes (van Hooijdonk en Kraemer, 2006). Na het bestuderen van de informatie moesten de proefpersonen de oefeningen ook daadwerkelijk uitvoeren. Daarbij bleek tekst het minst efficiënte medium te zijn: wanneer de oefeningen in de vorm van een tekst werden beschreven, hadden de proefpersonen de meeste tijd nodig om de informatie te bestuderen en de oefeningen uit te voeren. Afbeeldingen bleken het meest efficiënt. Bij moeilijkere oefeningen bleken filmpjes het meest effectief: de proefpersonen die een filmpje hadden gezien, maakten de minste fouten bij het uitvoeren. Bij RSI-oefeningen blijkt een afbeelding dus inderdaad meer te zeggen dan woorden, en ook minstens even goed te zijn als een filmpje voor het overbrengen van informatie. Als de proefpersonen het voor het kiezen hadden, gaven ze echter de voorkeur aan een filmpje.

Helaas hebben in de praktijk meestal systeem noch gebruiker iets te kiezen: veelal is een antwoord alleen in tekstuele vorm beschikbaar. Om toch visuele informatie te kunnen leveren, is in IMOGEN een methode ontwikkeld om uit een verzameling afbeeldingen een passende illustratie bij een tekst te zoeken (Bosma 2005b). Elke afbeelding in de verzameling is gekoppeld aan een stuk tekst: de tekst waar de afbeelding oorspronkelijk een illustratie van was (de tekst-afbeelding paren zijn afkomstig van Internet). Wanneer nu een antwoord door het QA systeem voorzien dient te worden van een afbeelding, wordt in de verzameling paren

van tekst + afbeelding gezocht naar een tekst die inhoudelijke overeenkomst vertoont met het antwoord. Om deze overeenkomst te bepalen wordt een beproefde, op statistiek gebaseerde techniek gebruikt die Latent Semantic Analysis (LSA) heet. Wanneer een geschikte tekst wordt gevonden, wordt de bijbehorende afbeelding als illustratie aan het antwoord toegevoegd. Zie uitbreiding (2) in figuur 1 voor een grafische weergave van dit proces. Het achterliggende idee is dat een afbeelding die bij de ene tekst past, ook bij een andere vergelijkbare tekst zal passen. Informele observaties wijzen uit dat deze methode inderdaad vaak relevante, maar soms ook minder voor de hand liggende illustraties oplevert. Verdere afstemming en evaluatie zijn dus zeker nodig.

Een laatste vorm van visuele informatie waar binnen IMOGEN onderzoek naar wordt gedaan, is informatiepresentatie door middel van een geanimeerd 'sprekend hoofd' dat als personificatie van het QA systeem fungeert, en de antwoorden met behulp van Nederlandse synthetische spraak aan de gebruiker overbrengt. Dit sprekende hoofd is ontwikkeld aan de Rutgers University in de USA¹; in het IMOGEN project is er Nederlandstalige spraaksynthese aan toegevoegd.² Een belangrijk aandachtspunt bij dit onderzoek is de rol die visuele signalen zoals wenkbrauwbewegingen of kijkrichting spelen in combinatie met spraak bij het op natuurlijke wijze overbrengen van informatie.

Conclusie

Het IMOGEN project is ongeveer halverwege en heeft tot nu toe veelbelovende resultaten opgeleverd op het gebied van zinsfusering, antwoordextensie en multimodaliteit, waarvan de meeste zijn toegepast in het QA-demonstratiesysteem dat binnen IMIX is ontwikkeld. Deze resultaten zijn echter niet alleen relevant voor QA, maar ook voor andere taaltechnologische toepassingen zoals multimodale informatiepresentatie en automatisch samenvatten. In de context van QA systemen is onze hypothese dat de binnen IMOGEN ontwikkelde technieken er toe leiden dat antwoorden relevanter en informatiever worden, en beter verifieerbaar zijn ("Is dit een antwoord op mijn vraag?"). De komende jaren zullen, naast het verfijnen en integreren van de ontwikkelde technieken, dan ook besteed worden aan het toetsen van deze hypothese.

Referenties

Bosma, W. (2005a), Extending answers using discourse structure. Proceedings of the RANLP Workshop on Crossing Barriers in Text Summarization Research, 24 september 2005, Borovets, Bulgarije.

Bosma, W. (2005b), Image retrieval supports multimedia authoring. Proceedings of the ICMI Workshop on Multimodal Interaction for the visualization and exploration of scientific data, 3 oktober 2005, Trento, Italië.

Hooijdonk, C.M.J., van en Kraemer, E. (2006), Information modalities for procedural instructions: the influence of text, static and dynamic visuals on learning and executing RSI exercises, ingediend bij Learning and Instruction.

Marsi, E. en Kraemer, E. (2005a), Explorations in Sentence Fusion. Proceedings of the 10th European Workshop on Natural Language Generation, 8-10 augustus 2005, Aberdeen, Schotland.

Marsi, E. en Kraemer, E. (2005b), Semantic classification by humans and machines. Proceedings of the ACL 2005 Workshop on Empirical Modeling of Semantic Equivalence and Entailment, 20 juni 2005, Ann Arbor, Michigan.

¹ Zie <http://www.cs.rutgers.edu/~village/ruth/>.

² Zie <http://nextens.uvt.nl/>. Een demo is beschikbaar op deze website.