

Ook registers van Handelingen ontsloten

Massadigitalisering is monnikenwerk

Na de lancering van de website www.statengeneraaldigitaal.nl op 8 juni is het project voor de digitalisering van de Handelingen van de Staten-Generaal in volle gang. In juni waren nog alleen de jaargangen 1990-1995 in te zien en te doorzoeken. Maar in 2010 zullen de verslagen en vergaderstukken van de Eerste en de Tweede Kamer uit de hele periode van 1814 tot 1995 (2,3 miljoen pagina's) gedigitaliseerd en online zijn.

In het project Staten-Generaal Digitaal 1814-1995 werken de Tweede Kamer en de Koninklijke Bibliotheek samen. Doel is het conserveren en digitaal ontsluiten van de Handelingen. Behalve omvangrijk zijn de Handelingen ook complex en moeilijk toegankelijk, zelfs voor ervaren onderzoekers. Digitalisering maakt het mogelijk om via een website rechtstreeks in het materiaal te zoeken op onder andere vrije zoekterm, naam, jaartal, datum en documenttype.

Afgelopen zomer is een begin gemaakt met het ontsluiten van de registers van de Handelingen op de website. In de periode 1814-1979 zijn bij de meeste jaargangen registers verschenen. Vanaf 1979 hebben beide Kamers een geautomatiseerd zoekstelsel ontwikkeld. De registers bestaan uit alfabetische lijsten van begrippen (de zaakregisters) en personen (de persoonsregisters) met daarachter verwijzingen naar relevante pagina's in de Handelingen.

De registerpagina's worden eerst als afbeelding ingescand. Het tonen van een afbeelding alleen geeft echter geen meerwaarde boven het origineel. Het wordt pas interessant als je via het register direct door kunt linken naar de passage in de Handelingen die je zoekt. Daartoe moet het origineel of de afbeelding omgezet worden naar computerleesbare tekst. De huidige OCR-software voor automatische tekenherkenning kan niet goed met historische lettertypes en taal omgaan en laat veel stekken vallen, terwijl in de beknopte stijl van de registers elke letter, elk cijfer en elk leesteken telt. Volledig overtypen of handmatige correctie zal daarom voor de digitale ontsluiting noodzakelijk zijn. Bovendien is de opzet, lay-out en de methode van verwijzen niet consistent over de 165 jaar die de registers beslaan. Dat betekent dat deze informatie niet geautomatiseerd verwerkt kan worden.

Onder één lemma vallen vaak tientallen pagina's met verwijzingen, die vooral bestaan uit afkortingen gevolgd door nummers van documenten, pagina's en data. Het zijn eindeloze, bijna abstracte opsommingen van letters, Romeinse en Arabische cijfers en leestekens. Om links te kunnen creëren moeten alle verwijzingen handmatig in een database ingevoerd worden. En dat is geen domme data-invoer: de datatypisten moet afdoende geïnstrueerd worden zodat zij begrijpen waar zij welke informatie moeten invoeren. De kleinste vergissing zorgt voor een niet werkende link.

Al dit werk gaat resulteren in een aparte zoekpagina op de website www.statengeneraaldigitaal.nl. De gebruiker kan kiezen uit een lijst van alle lemma's en deze in een zoekvraag combineren met een jaar of een periode. Als resultaat krijg je een lijst met alle vermeldingen van dat lemma in de registers van de betreffende jaren. Je kunt elke vermelding aanklikken, waarna je alle verwijzingen op je scherm krijgt naar passages in de Handelingen die op dit lemma betrekking hebben. Van elke verwijzing kun je rechtstreeks naar de desbetreffende pagina doorlinken. Door heen en weer te navigeren tussen de lijst met verwijzingen en de documenten is het mogelijk om in relatief zeer korte tijd met behulp van de registers de Handelingen grondig te doorzoeken.

Aan de digitale ontsluiting van historische bronnen gaat dus een enorme hoeveelheid monnikenwerk vooraf. Het spanningsveld tussen kwaliteit (nauwkeurigheid en volledigheid) en kwantiteit (2,3 miljoen pagina's in totaal, 17 duizend pagina's registers) is hierin een voortdurend aandachtspunt.