
Column

Charles van den Heuvel

The In-Between Machine

Het World Wide Web wordt een steeds belangrijkere bron voor onderzoek. Tim Berners-Lee beschreef in *Weaving the Web* (1999) zijn droom dat het web niet alleen een ruimte zou zijn om informatie te zoeken, maar ook om kennis aan toe te voegen. Uit het succes van FaceBook, MyTube, Flickr, maar ook van wetenschappelijke collaboraties, blijkt dat dit nu ook gebeurt. Tegelijkertijd verdwijnen er dagelijks miljoenen pagina's. In 1996 besloot Brewster Kahle het Web te archiveren. Vele nationale bibliotheken, waaronder de KB, volgden zijn voorbeeld met het archiveren van websites met nationale domeinnamen. Daarmee zijn deze webarchieven echter nog niet toegankelijk voor gebruik. Het blijft een vreemde gedachte dat je, om webarchieven van de Library of Congress te raadplegen, naar Washington moet afreizen om ze op monitors binnen de muren van de bibliotheek te lezen. Behalve problemen in verband met auteursrecht en nationaal collectiebeleid, zijn er andere redenen waarom webarchieven niet optimaal voor onderzoek kunnen worden benut. Allereerst zijn bestaande interfaces gebrekkig. Verder zijn de tools om websites door middel van annotaties in context te plaatsen, meestal voor individueel gebruik. Hierdoor is het voor toekomstige onderzoekers niet duidelijk of, en hoe, deze websites eerder zijn bestudeerd.

Wie nu webarchieven wil bekijken in Brewster Kahle's InternetArchive moet een url intypen in de Wayback Machine om vervolgens een lange rij van data te doorlopen waarop de sites zijn gearchiveerd. Dat is handig voor degene die meer te weten wil komen over 9/11 of een overzicht wil hebben van een betrekkelijk korte periode van intensief gebruik van websites, zoals die van de politieke verkiezingen in de VS. Echter voor de meeste vormen van historisch onderzoek is een dergelijke opsomming van data volstrekt ongeschikt. Daarom hebben we allereerst behoefte aan een betere interface. De ontwikkeling hiervan is geen eenvoudige zaak. Kies je voor een generieke, of juist voor een op specifieke disciplines afgestemde 'tailor-made' interface? Deze vraag lijkt onoplosbaar; zeker wanneer je bedenkt dat gebruikers kunnen wisselen en tijdens zoekopdrachten veranderen van strategie. Misschien moeten we ons daarom minder op de inhoud en meer op de stijlen van onderzoek richten, die onderzoekers gemeen hebben.

Onderzoekers moeten niet alleen relevante informatie vinden, maar deze ook met aantekeningen kunnen verrijken. Nu bestaan er diverse annotatietools voor particulier gebruik. Kennisverrijking van webarchieven maakt echter ook deel uit van ons collectief geheugen. Het zou zelfs een selectiecriteria van behoud kunnen zijn. Welk exemplaar van Newton's Principia zou jij behouden, die met of zonder de annotaties van Einstein? Samen met een aantal nationale en internationale instellingen werkt de Virtual Knowledge Studio for the Humanities and Social Sciences (KNAW) aan een aanvraag voor een Europese subsidie voor de ontwikkeling van de 'In-between Machine', waarbij de interface op onderzoek wordt afgestemd en webarchieven en annotaties als collectief geheugen worden bewaard.

Charles van den Heuvel is senior researcher bij de Virtual Knowledge Studio