

Slim en efficiënt zoeken met AGNES

Graven in archeologische onderzoeksrapporten

Archeologen in Nederland produceren zo'n 4.000 opgravingsrapporten per jaar. Alex Brandsen onderzoekt hoe deze schat aan informatie beter ontgonnen kan worden.

Steven Claeysens

Nederlandse archeologen hebben op dit moment zo'n 60.000 rapporten digitaal beschikbaar, bij DANS en in andere e-depots. Al deze rapporten samen bevatten een gigantische hoeveelheid archeologische informatie, maar het is heel moeilijk om hierin

alle relevante informatie over een bepaalde plaats of periode terug te vinden. De huidige systemen doorzoeken namelijk alleen de metadata van de rapporten. Deze metadata beschrijven bijvoorbeeld dat een rapport de Middeleeuwen behandelt, maar vermelden niet dat er ook enkele artefacten uit de Bronstijd zijn gevonden, terwijl deze objecten belangrijk zouden kunnen zijn voor

een onderzoek over de Bronstijd. Daarom is het nodig om alle tekst goed doorzoekbaar te maken. Alex Brandsen nam deze taak op zich als promovendus aan de Universiteit Leiden.

Taal begrijpen

Na een archeologie-bachelor in Leiden, een master Archeological Information Systems in York en ervaring als web developer in Leeds, startte Brandsen in 2017 met zijn promotieonderzoek. Hij wil de Ne-

bleem komt voor, namelijk wanneer één woord verschillende betekenissen heeft. Om al deze complicaties het hoofd te bieden, moet een zoekstelsel taal tot op zekere hoogte 'begrijpen' en ook specifiek archeologische concepten kunnen herkennen."

AGNES

"In mijn project pas ik text mining (en specifiek Named Entity Recognition) toe om automatisch relevante archeologische concepten te herkennen in tekst. Hiervoor gebruik ik machine learning, een vorm van kunstmatige intelligentie die op basis van voorbeelden uit handmatig geannoteerde teksten nieuwe woorden automatisch kan classificeren. In het verleden is daar mee geëxperimenteerd, een bruikbaar systeem heeft het helaas nog niet opgeleverd. Het doel van mijn project is om een webapplicatie te bouwen: AGNES (Archeological Grey literature Named

Entity Search). Met AGNES zoeken archeologen op een slimme en efficiënte manier door die stapels Nederlandse opgravingsrapporten, waardoor sneller en beter onderzoek te verrichten is in de Nederlandse

JONG TALENT

derlandse archeologische rapportproductie veel dieper ontsluiten: "Dat kan met full text-zoeken, zoals in Google, maar ook dan kunnen zich problemen voordoen. Bij de zoekterm Middeleeuwen vindt een full text-zoekactie bijvoorbeeld niet 'Middeleeuwse' en zeker niet '1000 na Christus'. Deze synonimie is een veelvoorkomend fenomeen in rapporten. Ook het omgekeerde pro-



Brandsen past textmining toe voor het automatisch herkennen van relevante archeologische concepten in teksten

archeologie." Een aantal versies van AGNES staan al online en kunnen door iedereen na registratie gebruikt worden.

agnesearch.nl