



Universiteit
Leiden
The Netherlands

Digital thesauri as semantic treasure troves: a Linguistic Linked Data approach to "A Thesaurus of Old English"

Stolk, S.S.

Citation

Stolk, S. S. (2023, May 31). *Digital thesauri as semantic treasure troves: a Linguistic Linked Data approach to "A Thesaurus of Old English"*. Retrieved from <https://hdl.handle.net/1887/3619351>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3619351>

Note: To cite this publication please use the final published version (if applicable).

Samenvatting (Dutch summary)

Thesauri zijn waardevolle informatiebronnen binnen de wetenschap. Door hun onomasiologische ordening, van betekenis naar woorden welke die betekenis uitdrukken, zijn deze lexicografische bronnen waardevol voor taalkundig en cultureel onderzoek. De eenvoud waarmee zulk onderzoek uitgevoerd kan worden, is afhankelijk van de vorm waarin een thesaurus beschikbaar is gesteld. Om die reden verkent deze dissertatie een nieuwe publicatievorm voor thesauri, specifiek gericht op zulke werken die historische taalvarianten beslaan, waarbij wordt getoetst of de vorm additionele gebruikersfunctionaliteit kan faciliteren. De digitale vorm die wordt getoetst is Linguistic Linked Data, die mechanismen van het Web gebruikt om taalkundige gegevens aan elkaar te koppelen en daarmee een netwerk van informatie vormt dat kan worden bevraagd.

Het eerste deel van de dissertatie maakt zowel bestaande als gewenste karakteristieken inzichtelijk van thesauri van historische talen. Hoofdstuk 1 benoemt daartoe de elementaire bouwstenen waaruit de inhoud van thesauri bestaat. Dit overzicht berust zich op een analyse van beschikbare thesauri voor historische varianten van het Schots en het Engels. Een thesaurus, als zijnde een semantisch georganiseerd woordenboek, bestaat uit drie hoofdbouwstenen: (1) een hiërarchie van semantische begrippen; (2) woorden of frasen in een bepaalde betekenis, gepositioneerd binnen de eerdergenoemde hiërarchie; en, optioneel, (3) relaties van synonymie tussen woorden of frasen in een dergelijke betekenis. De daadwerkelijke informatie die is vastgelegd middels deze bouwstenen verschilt van thesaurus tot thesaurus. Zoals Hoofdstuk 2 aantoont, wensen onderzoekers de inhoud van thesauri te benutten als opstap voor het verkennen van gerelateerde materie die buiten het kader valt van het gebruik zoals initieel beoogd door de samenstellers van de thesaurus. De mogelijkheid om inhoud van thesauri te hergebruiken, uit te breiden en er naderhand onomasiologische analyses op te verrichten, zijn de noemenswaardigste nieuwe functionaliteiten die onderzoekers wensen en die vooralsnog ontbreken in gepubliceerde thesauri van de historische varianten van het Schots en het Engels.

Volgend op de analyse van de inhoud en functionaliteit gewenst voor thesauri van historische talen, beschouwen Hoofdstukken 3 tot en met 5 in welke digitale vorm deze lexicografische werken gepubliceerd zouden moeten worden op het Web. De voorgestelde vorm is gebaseerd op standaarden van het Semantic Web, met name SKOS en Lemon-OntoLex, en wordt recentelijk aangeduid als Linguistic Linked Data. De huidige specificaties van deze twee standaarden bleken niet afdoende om thesauri volledig vast te leggen; mijn werk rondom de ontwikkeling van het *lemon-tree* model was erop gericht om dit hiaat te vullen. Naast het aangeven op welke wijze de twee eerdergenoemde standaarden gecombineerd dienen te worden voor het vastleggen van thesauri, dekt *lemon-tree* twee verdere aspecten van thesaurusinhoud die van belang zijn voor dit doel: (1) niveaus die te onderscheiden zijn binnen de hiërarchieën van semantische concepten en (2) een lossere vorm van categorisering dan lexicalisatie. Deze toevoegingen zijn niet enkel relevant voor thesauri van historische talen, maar

tevens voor thesauri in het algemeen.

De laatste vier hoofdstukken evalueren het nut van de Linguistic Linked Data-vorm van thesauri door deze toe te passen op *A Thesaurus of Old English* (*TOE*), een thesaurus die de vroegmiddeleeuwse variant van het Engels beslaat. Het resultaat van de transformatie van *TOE* van het oorspronkelijke formaat naar Linguistic Linked Data (ook wel *TOE*-LLD genoemd), zoals besproken in Hoofdstuk 6, is beschikbaar gesteld in de webapplicatie Evoke. Deze applicatie is specifiek ontwikkeld voor thesauri als onderdeel van dit doctoraal onderzoek en wordt behandeld in Hoofdstuk 7. Evoke en *TOE*-LLD zijn gebruikt in verscheidene casussen binnen het onderzoeksproject ‘Exploring Early Medieval English Eloquence’ (EEMEE). Hoofdstuk 8 biedt een overzicht van deze casussen, inclusief reflectie op de waarde van de twee bronnen die erin centraal staan. Uit de evaluatie van de gekozen digitale vorm voor thesauri en de functionaliteit die Evoke biedt, komt naar voren dat de combinatie van de twee een krachtig middel kan zijn voor onderzoekers om een thesaurus met extra informatie te verrijken. Onderzoekers hebben aangetoond dat ze met deze middelen tot hun beschikking, vernieuwend onderzoek hebben kunnen doen naar, onder andere, historische ontwikkelingen binnen de lexicografie, stilistiek, diachrone taalontwikkelingen en verschillen tussen verwante historische talen.