


Leximancer 5.0

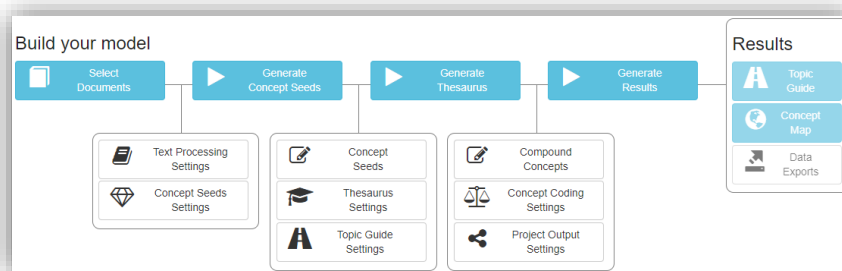
	Text Analysis	
	Systeemvereisten:	Cloud / Windows 7+ / OSX 10.11+
	Ontwikkeld door:	https://www.leximancer.com/
Commercieel	Contactpersoon:	joachim.ganseman@smals.be

Functionaliteiten

Leximancer is een analyseprogramma dat uit een grote hoeveelheid tekst de belangrijkste onderwerpen identificeert, deze groepeerd in gerelateerde concepten, en de relaties ertussen visualiseert. Het kan gebruikt worden om de grote lijnen van een lange tekst te identificeren, of de belangrijkste thema's in de resultaten van een bevraging. Leximancer ondersteunt een veelheid aan formaten waaronder MS Office, PDF, HTML, CSV etc. Er worden 18 talen ondersteund waaronder alle Belgische landstalen.

Leximancer biedt een gemakkelijk bruikbare grafische interface die de gebruiker bij elk nieuw project begeleidt door alle stappen. Bij elke stap kunnen verschillende parameters ingesteld worden om de analyse te verfijnen, of de resultaten manueel aan te passen. Het eindresultaat is een dynamische grafiek waarvan sommige eigenschappen, zoals grootte van clusters en hoeveelheid van clusters, ook met sliders kunnen worden aangepast. Grafische output kan worden bewaard als afbeelding. De resultaten van de analyse kunnen ook als spreadsheet gedownload worden voor verdere verwerking in andere analyseprogramma's.

Goede versionering ontbreekt, wat een gemis is bij grote projecten die veel rekentijd vragen.



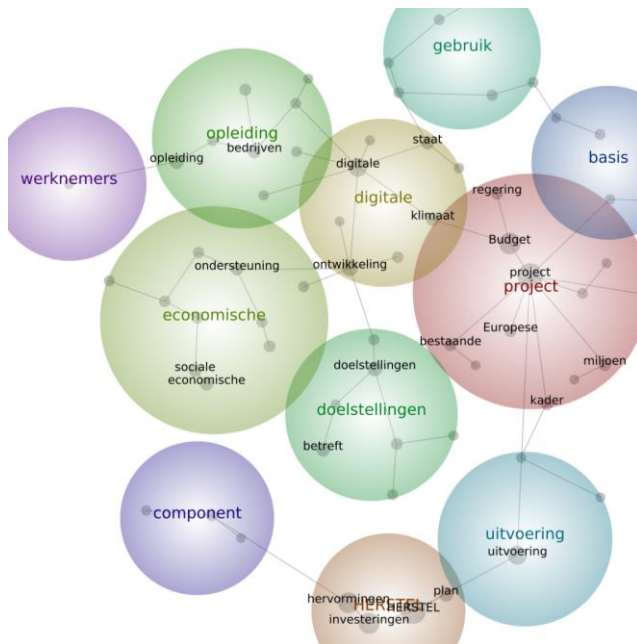
Conclusies & Aanbevelingen

Wie regelmatig geconfronteerd wordt met grote hoeveelheden onbekende tekst, heeft met Leximancer een workflow voorhanden om structuur in de gegevens te ontdekken en inhoudelijk een rode draad te vinden. De input van de gebruiker blijft noodzakelijk om tot betekenisvolle resultaten te komen. Voor optimaal gebruik moet de handleiding van 150 pagina's gelezen worden. Leximancer is een tool voor "e-discovery", en mogelijke toepassingsdomeinen zijn onder andere inspecties, forensisch onderzoek, grootschalige bevragingen of snelle literatuurreviews.

Testen & Resultaten

Het “nationaal plan voor relance en veerkracht” is een PDF document van bijna 700 blz. De inhoudstabel organiseert het op thematiek (mobiliteit, klimaat, economie, samenleving, digitaal etc.) maar misschien zijn we eerder geïnteresseerd in alle aspecten doorheen de tekst die te maken hebben met een bepaald beleidsniveau (federaal/regionaal/lokaal) of een bepaalde sector (chemie, agro, ...).

Na upload van de tekst en instellen van de taal moeten we parameters instellen voor de analyse, zoals de hoeveelheid tekst die per keer in acht genomen moet worden of het aantal clusters en de grootte ervan. Er zijn verschillende opties beschikbaar om bijvoorbeeld duplicaten te negeren (o.a. voor datasets waarin veel wordt geciteerd, zoals emails) of veel gebruikte woorden (“de”, “het”, “en”, “of”, ...) over te slaan. Manuele tussenkomst is altijd wenselijk om gevonden woorden te fuseren als 1 onderwerp, we geven o.a. aan dat “overheid” en “staat” als synoniemen beschouwd mogen worden, alsook “onderneming” en “bedrijf” en nog een tiental andere. Ook eigen concepten en vaste uitdrukkingen kunnen toegevoegd worden.



Een grafische conceptenkaart (zie hiernaast) geeft tot slot een algemeen overzicht van de tekst. We kunnen erin opmerken dat “werknemers” aan “bedrijven” gelinkt zijn via “opleiding” in deze tekst. Verder staat het digitale erg centraal, en ook het sociale en economische komen vaak in elkaars buurt voor.

Daarnaast gebeurt er een analyse van de “onderwerpen”, concepten die regelmatig samen voorkomen, op basis waarvan de tekst doorzocht kan worden. Zo is er “Behoeften + Digitale”, waarbij men 175 paragrafen vindt die min of meer te maken hebben met digitalisering en digitale (on)geletterdheid. Over “Gebouwen + Energie” zijn er 152 paragrafen over o.a. renovatie en energie-efficiëntie van infrastructuur.

Een grotere dataset van ongeveer 9000 arresten van het Hof van Cassatie, vergt een tiental minuten analysetijd, te herhalen bij elke grote aanpassing aan de concepten. Hetzelfde stappenplan volgend, vinden we na wegfiltering van de meeste juridische concepten, dat de meeste tekst daarin handelt over (geschillen met betrekking tot) arbeidsovereenkomsten, verkeersongevallen, en verkoop van goederen.

Gebruiksvoorwaarden & Budget

Leximancer is beschikbaar als cloud-oplossing met een maandelijks of jaarlijks abonnement, of als standalone desktop programma met een jaarlijks hernieuwbare licentie. Voor academisch en wetenschappelijk onderzoek zijn er kortingen. Een gratis trial van 30 dagen is mogelijk. De recentste prijzen staan op <https://www.leximancer.com/desktop-portal>