

ZONDER BEGRIPPENWOORDEN BOEK GEEN BIG DATA

Bij veel organisaties ontstaat een wirwar aan woordenlijsten

De standaard **Simple Knowledge Organization System** is sinds kort verplicht bij de overheid. De voordelen, zegt Jan Voskuil, zijn duidelijk, ook voor het aansturen van IT-projecten. Om te beginnen is een goed gestructureerd begrippenwoordenboek de basis voor kennismanagement. En deze standaard is daar een solide basis voor. Bovendien ontstaan er meer mogelijkheden voor ontsluiting en kunnen de woordenboeken van verschillende organisaties worden gekoppeld.

door: JAN VOSKUIL beeld: PESHKOVA / SHUTTERSTOCK.COM



Sinds 18 mei 2015 staat SKOS (Simple Knowledge Organization System) op de 'pas toe of leg uit'-lijst van Bureau Forum Standaardisatie. Daarmee is het een verplichte standaard voor overheidsorganisaties. De standaard wordt veel gebruikt bij het ontsluiten van grote kennisbanken en webbibliotheken. Brede toepassing van de standaard kan een impuls geven aan de kwaliteit die gerealiseerd wordt bij het vastleggen en ontsluiten van het bedrijfsvocabulaire. Deze van sport, in het Engels aangeduid met Enterprise Vocabulary Management, wordt vanouds beoefend in grote, wereldwijd opererende concerns. In toenemende mate zien ook kleinere organisaties de grote voordelen die deze discipline met zich meebrengt.

Elke organisatie hanteert begrippen in de specifieke betekenis. Zo is in het gevangeniswezen de 'badmeester' iemand die nieuwe 'inkomsten' behandelt: inname eigendommen, uitrekenen beddengoed et cetera. De Belastingdienst hanteert uiteraard een heel andere definitie van 'inkomsten'. Kennis van dergelijke organisatie-

specifieke begrippen is essentieel. Het is dan ook een goed idee om deze kennis structureel en systematisch vast te leggen in een begrippenwoordenboek en te ontsluiten op een doeltreffende manier.

Voordelen

De voordelen die dat biedt zijn in te delen in zes thema's. In de eerste plaats vormt een goed gestructureerd begrippenwoordenboek de basis voor kennismanagement. Medewerkers die willen nazoeken wat een specifiek begrip ook al weer precies betekent, worden zo doeltreffend bediend.

Directe beschikbaarheid van goed doordachte definities verhoogt daarnaast de kwaliteit van de informatievoorziening.

Een derde thema is beheersbaarheid van informatie. Veel organisaties hebben te maken met een veelheid van informatiesystemen. Het houden van overzicht over alle attributenameningen in de vaak talloze onderliggende databases is een grote uitdaging. Dat vraagt om een structurele aanpak.

Een professionele aanpak bij het vastleggen van het bedrijfsvocabulaire biedt ook voordelen op het vlak van vindbaarheid van informatie en bruikbaarheid van gegevens. Het toekennen van trefwoorden aan artikelen verhoogt de precisie waarmee experts een kennisbank kunnen doorzoeken. Goede, verklarende definities die eenvoudig toegankelijk zijn verhogen de waarde van zulke trefwoordsystemen. Hetzelfde geldt natuurlijk ook voor managementinformatie en business intelligence. Wanneer het aantal KPI's klein is, volstaat een simpel lijstje definities. Naarmate er meer en grotere gegevensbestanden worden meegenomen, neemt het aantal gehanteerde begrippen snel toe. En daarmee ook het belang van definities: zonder begrippenwoordenboek geen Big Data. Een vijfde thema is interoperabiliteit. Daarbij denk je in de eerste plaats aan het uitwisselen van informatie tussen IT-systemen. Een goed afgestemd vocabulaire is daarvoor een absolute randvoorwaarde. Vocabularymanagement draagt echter ook bij aan uitwisselbaarheid van informatie op intermenselijk niveau. In de na-

lijsten

HOE WERKT SKOS?

Het basisidee van SKOS is dat elk begrip een eigen webadres krijgt. Een fictief voorbeeld is <http://bestelingsdienst.nl/begrippen/Inkomsten>. Het begrip 'inkomsten' zoals het gevangeniswezen dat hanteert krijgt weer een ander webadres. Aan het webadres van een begrip is meteen te zien welke organisatie verantwoordelijk is voor de definitie van het begrip. Dat is niet alleen simpel en doeltreffend, het geeft ook eenduidigheid op een ongekende schaal. Het gebruik van SKOS zorgt er niet alleen voor dat de identiteit van elke begrip in de thesaurus ondubbelzinnig vaststaat en de betekenis ervan eenvoudig is op te zoeken. SKOS voorziet ook in middeelen om begrippen in de ene thesaurus te linken aan begrippen in een andere thesaurus. Dat gaat volgens een standaardrecept. Voor organisaties die met elkaar samenwerken is het handig om elkaars definities altijd bij de hand te hebben, ook over landsgrenzen heen. De Europese Unie (EU) heeft een aantal belangrijke thesauri laten maken op basis van SKOS met begrippen in 23 EU-talen. Overheidsorganisaties kunnen de begrippen in hun eigen thesaurus eenvoudig linken aan begrippen in deze EU-thesauri. Dat bevordert het delen van informatie op Europees niveau.

sloop van de zaak Dolmatov ontstond er discussie over het exacte aantal suicides in inrichtingen waar mensen worden vastgezet: DJ1 en IGZ kwamen met andere aantallen. Dat hoeft natuurlijk niet veroorzaakt te worden doordat deze organisaties verschillende definities hanteren. Maar als geïnteresseerde burger zou je dat eenvoudig willen kunnen nazoeken. Ten slotte is een professionele aanpak van vocabulairebeheer van groot belang bij het aansturen van IT-projecten. In het ideale geval kan elk IT-project binnen de organisatie putten uit een breed gedragen, actueel begrippenwoordenboek. Als het project nieuwe werkprocessen en, daarmee samenhangend, nieuw begrippen introduceert, dan is er een staande procedure waarmee de nieuwe begrippen aan het woordenboek worden toegevoegd. Dat klinkt logisch, maar de praktijk is helaas meestal anders.

Belang van governance

De praktijk bij veel organisaties is dat het woordenboek er helemaal niet is, of een lage dekkingsgraad heeft van slechts een handjevol begrippen. IT-projecten gaan dan noodgedwongen zelf aan de slag met het maken van definitielijsten. Zo ontstaat er in de loop van de tijd een wirwar aan woordenlijsten, die via mailwisselingen in diverse versies rondgaan in de organisatie. Soms levert een IT-project een gegevenswoordenboek op. Eigenaarschap ligt dan meestal bij de technisch beheerder en de inhoud is voor gebruik in de business te gedetailleerd en te technisch. Nieuwe begrippen die het IT-project introduceert, worden binnen het project bedacht en gedefinieerd. De definities zijn dan te vinden in de ontwerpdocumentatie. Op die manier ontstaat verwarring en dat doet afbreuk aan de geleverde kwaliteit.

Het vastleggen van het bedrijfsvocabulaire in een begrippenwoordenboek is dus van groot belang. Door het begrippenwoordenboek te ontsluiten op het web ontstaat een 'single source of truth'. Essentieel daarbij is het inrichten van een effectieve governancestructuur. Wat in de

praktijk goed werkt is een groep te vinden die binnen de organisatie op tactisch-strategisch niveau over de bedrijfsvoering gaat. Wijzigingsvoorstellen op het begrippenwoordenboek worden binnen deze groep besproken en goedgekeurd. Periodiek, bijvoorbeeld eens per jaar, worden alle wijzigingen van de afgelopen periode voorgelegd aan een overleg op het hoogste niveau binnen de organisatie. Zo ontstaat een 'release-cyclus' waarin het woordenboek op gecontroleerde wijze up-to-date wordt gehouden. Tegelijkertijd heeft het woordenboek een duidelijke statuur, wat een einde maakt aan oeverloze discussies. Het eigenaarschap ligt duidelijk bij de business.

Zo weten IT-projecten waar ze aan toe zijn. Het belang van een adequaat begrippenwoordenboek en de processen eromheen gaat echter veel verder dan alleen het aansturen van IT-projecten, zoals we hierboven hebben gezien.

Inhoud en structuur

Als de governance is ingericht moet het woordenboek opgesteld worden. Een eerste vraag die zich daarbij voordoet is de gewenste dekkingsgraad. Een te hoog ambitieniveau werkt verlamdend. Belangrijk is ook goed nota te nemen van wat lexicografen sinds jaar en dag weten: de frequentieverdeling van woorden neemt exponentieel af. Onderzoek toont dat je met 5000 Nederlandse woorden 95 procent van alle teksten kunt begrijpen. Voor de overige 5 procent is kennis vereist van nog eens 200.000 woorden. Dat betekent dat je met een relatief beperkt aantal trefwoorden potentieel al een zeer effectief begrippenwoordenboek in handen hebt. Aan de andere kant moet het ambitieniveau ook niet te laag liggen. Het moet wel de bedoeling zijn om bijvoorbeeld IT-projecten effectief aan te kunnen sturen.

Bij het vastleggen van het begrippenwoordenboek is toepassing van de open standaard SKOS een goed idee. Het geeft een solide basis voor de structuur, het zorgt voor geavanceerde mogelijkheden voor ontsluiting en maakt het mogelijk om begrippenwoordenboeken van verschil-

lende organisaties aan elkaar te koppelen. Een woordenboek is soms niet meer dan een alfabetisch gesorteerde woordenlijst. Een hiërarchische ordening werkt veel beter. Om de betekenis van 'lente' te doordringen is het essentieel te weten dat het een nauwer begrip is van 'seizoen', naast 'zomer', 'herfst' en 'winter' – de SKOS-standaard noemt dit 'narrower'. Een woordenboek dat zulke semantische relaties expliciet beschrijft noemen we een thesaurus. SKOS biedt een gelimiteerd aantal standaardrelaties. Het laat de betekenis van de relaties doelbewust vaag, waardoor een SKOS-thesaurus een wat informeel karakter krijgt. Daarin schuilt veel van de kracht van deze standaard: de beschrijving is intuïtief, direct toegankelijk voor een breed publiek en biedt tegelijkertijd aantoonbaar meerwaarde. En omdat SKOS een webstandaard is, bieden de semantische relaties meteen ook een navigatiestructuur.

Web van begrippen

SKOS maakt het daarom mogelijk om woordenboeken die op deze manier gestructureerd zijn (thesauri) via het web toegankelijk te maken in een 'Wikipedia-achtige' vorm, zodat elk trefwoord een eigen artikel krijgt met definitie, toelichting en links naar gerelateerde trefwoorden. Zo kun je bij het artikel over 'seizoen' eenvoudig doorklikken naar de 'narrowers' (lente, zomer, herfst, winter).

SKOS ontsluit kennis over begrippen niet alleen via de browser, het zorgt er tegelijkertijd voor dat computerprogramma's de diverse relaties in de thesaurus slim kunnen gebruiken voor allerlei doeleinden. Dat komt doordat SKOS een lid is van de snel groeiende familie van Linked Data-standaarden, die het mogelijk maken naast documenten ook gegevens op het web te ontsluiten. Daar zit de crux van SKOS. Sommige systemen voor contentmanagement bieden bijvoorbeeld al plug-ins waarmee je op een webpagina alle woorden die voorkomen in de bedrijfsthesaurus, automatisch kunt voorzien van een hyperlink naar de begripsdefinitie. Eigenlijk zou elke organisatie zijn begrippenparaat op deze manier moeten publiceren. De fundamentele kennis die daarmee beschikbaar komt, vormt immers de basis voor de gehele informatiehuishouding, zowel intern als in de interactie met de buitenwereld. Hopelijk brengt het verplicht stellen van SKOS binnen de overheid dit ideaalbeeld dichterbij. <<



Jan Voskuil is directeur van Taxonic, een van de partijen die Platform Linked Data Nederland mogelijk maakt (jan.voskuil@taxonic.com).