

Het belang van de Thesaurus

Taalkundige verbanden

De verbanden die er bestaan tussen woorden onderling en tussen frases en woorden, is zeer bepalend voor het resultaat van zoeken en vragen beantwoorden. De taalkundige synoniemen worden centraal bijgehouden. Het Hylo synoniemenboek bevat bijna 100.000 geclassificeerde synoniemen. Deze classificatie maakt het mogelijk om het verschil tussen “bank” als meubelstuk of als financiële instelling uit elkaar te houden.

Maar ook worden door Hylo wederkerigheid en afgeleide verbanden onderscheiden. Zo is een “ezel” een dier en heeft in de klasse “dier” geen synoniem, wel als “domoor” of “sukkel” en als schilderobject. Ook de relatie tussen “baas”, “manager”, “chef” en “kok” is niet omkeerbaar.

Ditzelfde geldt tussen hyperoniemen en hyponiemen. Een “kalf” is een “koe”, is een “rund”, is een “beest”.

Andere verbanden

Behalve de taalkundige verbanden, die Hylo aanlevert, mag de Knowledge Manager zelf ook verbanden aanleggen. Verbanden die taalkundig mogelijk niet correct zijn, maar wel de taal van de klant is. Zo is een “abonnement” strikt genomen geen “contract”, maar zullen wel binnen een bepaalde context uitwisselbaar zijn. Hetzelfde geldt voor “opzeggen”, “beëindigen”, “stoppen”, “kappen”, “staken” en zo meer. Hylo zal, op basis van ervaring bij andere implementaties, via Hylo Exchange de Knowledge Manager suggesties doen om zulke verbanden ook op te nemen als de woorden in het Lexicon voorkomen. Dit bespaart de Knowledge Manager de tijd om hier zelf proefondervindelijk achter te komen.

Ook kan de Knowledge Manager groepsverbanden aanleggen, tussen woorden en begrippen. Zo kan een gebruiker die naar “een pan” zoekt 0 resultaten krijgen, of verwezen worden naar “keukengerei”. Dit fenomeen is ook omkeerbaar.

Naast verbanden tussen woorden, kunnen ook gezegdes en frases onderkend worden. Zo is “op de bus doen” binnen de context van post gelijk aan “versturen”.

Specifieke verbanden

Binnen de context van een organisatie, branche of vakgebied komen specifieke verbanden voor zoals hiervoor genoemd. Sommige kent de Knowledge Manager, andere worden aangereikt door Hylo Exchange en de overige leert u van uw gebruikers uit de vragen die zij stellen. De ervaring leert dat binnen ca. 6 weken actief gebruik, meer dan 80% van alle variëteiten in de vragen zijn voorgekomen.

Via de Resultaat Analyse (stoplichten) wordt de Knowledge Manager hierop gewezen en de gelegenheid geboden om nieuwe verbanden op te nemen. Ook kan de Knowledge Manager gebruik maken van de toolset om correspondentie via e-mail te analyseren.

Samenstellingen

Een bijzonder en in het Nederlands veelvoorkomend verschijnsel is de samenstelling van woorden. Een “reisverzekering”, “autoverzekering”, “levensverzekering” zijn alle samenstellingen. Wil men toestaan dat als men naar een “verzekering” vraagt, ook de samenstellingen worden gevonden, of juist niet? Hylo ondersteunt de Knowledge Manager met een samenstellinganalyse, waarbij de Knowledge Manager kan aangeven bij welke variaties van het leidende zelfstandig naamwoord wel of juist niet ook in verband gezien mogen worden.

Het gevolg hiervan is dat als een gebruiker zoekt naar “verzekering” wel of niet “reisverzekering” gevonden wordt. En, dat in de resultaatpresentatie de gebruiker eerst gevraagd kan worden welke verzekering bedoeld wordt: “reisverzekering”, “autoverzekering” of “levensverzekering”.

Toelichting

Voor het gebruik van een zoekmachine zijn de keuze en spelling van de woorden cruciaal. Immers, het gaat om de steekwoorden waarmee gezocht gaat worden. Technische gezien, wordt met een groepje letters gezocht, ongeacht wat de betekenis daarvan is. Het is niet alleen een probleem van de juiste spelling dat een rol speelt in het succesvol zoeken. Ook de keuze van de gebruikte woorden, synoniemen, aliassen, acroniemen, eigennamen, hyponiemen en hyperoniemen zijn bepalend voor het succes van de zoekopdracht. Een zoekmachine weet immers de relatie tussen vee, runderen, koeien, stieren, kalveren en vaarzen niet. Ziet zelfs niet de overeenkomst tussen kalf en kalfje. Ook de vaardigheid om expressies op te kunnen stellen voor zoekopdrachten en zoekverfijningen te kunnen formuleren als er te veel antwoorden gevonden worden is niet bij iedere internetgebruiker aanwezig. Dus zoeken naar antwoorden is daarom voor velen niet de passende methode om een antwoord te vinden.

Onze taal is veel krachtiger dan een set woorden, dus waarom niet gewoon klanten laten doen waar zij vertrouwd mee zijn, een vraag stellen in hun eigen taal? Als men een vraag stelt, dan krijgen alle woorden in de zin een betekenis. Als we willen weten hoe een “zakelijke rekening” te openen, heeft het natuurlijk geen zin als antwoord te geven hoe een “prive rekening”, een “internet spaarrekening” of een “beleggingsrekening” te openen. Of allerlei informatie met het woord “zakelijke” te presenteren. Om uw klanten het makkelijk te maken hun vraag te stellen, in hun eigen taal, is slimme taaltechnologie nodig, zodat wij klanten kunnen begrijpen. Tenslotte, als ze bellen of een mail sturen, nemen wij hun vraag ook serieus, ook al staan er spelfouten in, kiezen zij

Woordenboek 250.000 woorden

De basis voor Hylo is een uitgebreid woordenboek, met ruim 250.000 woorden. Het woordenboek speelt vooral een rol in de spelling en suggesties die aan vraagstellers gegeven kunnen worden wanneer zij fouten maken.



Woordenboek 15.000

Het woordgebruik van veel Nederlanders is beperkt. Met andere woorden, er is een groep van woorden die beduidend vaker gebruikt worden dan andere woorden. Deze zogenaamde frequente woorden zijn samengevat in een speciaal woordenboek met 15.000 frequent gebruikte woorden. Hylo gebruikt dit woordenboek voor de eerste controle uit efficiency oogpunt.

Jeugd woordenboek

De jongere generatie gebruikt een geheel eigen vocabulaire. Trendy woorden, vaak uit andere talen ontleend, msn taal en zelfs sms taalgebruik vinden wij terug bij de vraagstellingen. Om die toch te kunnen begrijpen worden deze “vertaald” naar woorden uit het standaardwoordenboek.

Persoonsamen, plaatsen en landen

Hylo's thesaurus beschikt ook over tabellen met de meest voorkomende persoonsnamen, de Nederlandse plaatsnamen en de meeste landennamen. De thesaurusbeheerder kan hieraan toevoegen.

lexicon

Naast de standaard woordenboeken bouwt Hylo een woordenboek van alle woorden die u gebruikt op uw website, uw knowledge base of ieder ander domein dat u aanwijst als bron om antwoorden in te vinden. Dit lexicon van al uw woorden is maatgevend voor wat wel en niet gezocht kan gaan worden, bepaalt de voorkeur spellingscorrectie en vult dynamisch de standaardwoordenboeken aan.

Aliassen

Om de relatie te kunnen leggen tussen standaardwoorden en specifieke woorden die gebruikt worden in de organisatie die de selfservice aanbiedt, kunnen aliaassen gedefinieerd worden.

Acroniemen en eigennamen

Ook de afkortingen van diensten en producten als mede de eigennamen van producten kunnen opgenomen worden, om klanten te helpen bij de formulering van de vraag en duidelijkheid te geven bij de semantische analyse.

Hyperoniemen en hyponiemen

Naast de vaste relatie tussen woorden als “rund”, “koe” en “kalf”, zijn er veelal producten en diensten die met verzamelnamen, product of type worden aangeduid. Zo kan binnen een bepaalde context “mobieltje”, “Nokia” en “6310” dezelfde betekenis hebben. Het is dan voor de aanbieder van de informatie niet nodig 3 versies van een antwoord te schrijven, alleen omdat de vraagsteller met verschillende woorden het zelfde bedoelt. Anderzijds, is het niet wenselijk dat bij een referentie naar een “6310” men geen antwoord krijgt. Want hoe kon de vraagsteller weten dat de antwoorden hetzelfde zijn voor alle “mobieltjes” en de aanbieder maar één antwoord heeft gemaakt.

Omgekeerd geldt dit ook. De vraagsteller die in een vraag naar “mobieltje” refereert, maar het antwoord afhankelijk is van het soort en type, dan is het niet wenselijk geen resultaten of een grote lijst van mogelijke antwoorden te presenteren. Beter is de relatie te onderkennen en de vraag te stellen: “over welk soort en type gaat het?” en dan een passend antwoord te geven.

Synoniemen

Alle standaard synoniemen van de Nederlandse taal zijn in de Hylo thesaurus opgenomen.

Spelling

Fuzzy spelling

Hylo gebruikt Fuzzy Logic technologie om bij spelfouten te bepalen welk woord waarschijnlijk bedoeld is. Hierbij spelen onder andere klank, veelvoorkomende typefouten (zoals omkeringen, naastliggende toetsen etc), score van de letters binnen het woord en percentage afwijking een rol.

Correctie suggesties

Bij keuze uit mogelijke correcties wordt het woord gekozen dat het meest waarschijnlijk is. Zo zal bij “nbiet” de suggestie “niet” in plaats van “bied”, “bid”, “biet” of “bit” gedaan worden als dat in context past. Bij de correctie speelt niet alleen de Fuzzy technologie een rol maar ook de thesaurus die tenslotte aangeeft welke woorden relevant zijn en welke niet.

Lange woorden

In de Nederlandse taal komen veel samengestelde woorden voor, terwijl op veel vakgebieden woorden voorkomen die uit andere talen zijn geleend of eigennamen zijn. Zodra deze woorden tijdens het typen herkend worden, worden reeds na de eerste letters al complete woordsuggesties aan de vraagsteller gedaan. Dit bespaart niet alleen tijd, maar geeft vraagstellers ook hertvertrouwen dat zij het goede woord te pakken hebben. Dit voorkomt dat bepaalde woorden omzeild worden door omschrijvingen die het resultaat negatief zouden kunnen beïnvloeden. Samengestelde woorden

Het maakt voor de vraagsteller niet uit of hij de woorden separaat schrijft, of als samenstelling.

Zo is motor rijbewijs gelijk aan motorrijbewijs. Als geen specifiek antwoord gevonden kan worden over motorrijbewijs, maar wel over rijbewijs, dan zal de vraagsteller hier op gewezen worden en de keuze gelaten worden of hij het antwoord al of niet wil inzien.

Woordkeuze en herleiding

De van de woorden is door de uitgebreide thesaurus met zijn woordverbanden vrij. Ook worden woorden “terug herleid naar de stam” en daarvan alle afleidingen betrokken in de analyse van de vraagstelling. Het maakt daarom voor de vraagsteller niet uit of hij grammaticale fouten maakt, bijvoorbeeld in het vervoegen van werkwoorden. Zo worden “zuigde stof” en “zoog stof” beiden terug herleid tot zuig, zuigen, gezogen etc.

Weten en niet weten

Wanneer vragen gesteld worden met woorden die geen raakvlak hebben met de geboden informatie, dan is er kans dat het antwoord niet gegeven kan worden. Dat is niet altijd wenselijk. Een en ander hangt af van de betekenis van het woord binnen de vraagstelling. Zo is de betekenis van een bijwoord of bijvoeglijk naamwoord minder relevant dan die van een zelfstandig naamwoord of werkwoord. Als het wel relevant is, dan moet de vraagsteller weten dat hij vragen stelt met woorden die niet te herleiden zijn naar de informatie die antwoorden kunnen bevatten. Daarmee wordt voorkomen dat de vraagsteller niet begrijpt waarom een

antwoord schuldig blijft. Ligt het aan de vraag, of ligt het aan de beschikbare informatie waaruit de antwoorden geput worden.

Hylo Exchange

Met het Hylo Exchange platform delen Knowledge Managers kennis met elkaar. Niet de informatie die specifiek voor u is, maar verbanden die van generiek belang zijn. Zo kunt u via Hylo Exchange opvragen welke generieke woordverbanden ander Knowledge Managers hebben toegevoegd aan de standaard Hylo Thesaurus. Het Hylo Exchange platform wisselt ook Topic en Issue mapsuggesties uit, waarmee ervaring gedeeld kan worden.

Thesaurus versus meta-data

Alle informatie in de Hylo.Thesaurus heeft een globale betekenis voor alle content. Dit in tegenstelling tot meta-data die aan de content wordt toegevoegd. De eigenschap daarvan geldt alleen voor de betreffende content en zal bij iedere mutatie weer bijgewerkt dienen te worden