



EasySeparate en Semantiek, een introductie met nieuwe technologie

Apeldoorn, 13 oktober 2009.

EasySeparate en Semantiek, een introductie

Semantiek betekent zoveel als de samenhang van woorden die een document vormen, oftewel: de betekenis van de tekst. Stel je voor dat binnenkomende documenten herkend kunnen worden op basis van hun *inhoud*.

Vandaag is dit een reële voorstelling van herkenningstechnologie.

Bij de installatie van EasySeparate is nu een voorbeeld beschikbaar om deze technologie te testen. Deze evaluatie laat onze nieuwe module op transparante wijze ontdekken met "nieuws-classificatie". Herkende nieuwsberichten komen in de uitvoer onder de gedefinieerde onderwerp-mappen te staan. Op dit moment is deze Semantiek-module getraind op Juridische correspondentie en nieuwsberichten.

Deze technologie gaat andere document structuren leren, onze factuurtemplate is hier een voorbeeld van.

Semantiek is de samenhang van woorden; zo is een offerte anders van toonzetting dan bijvoorbeeld een aanmaning. Met verschillende offertes en aanmaningen "leren" we de software, die uit een aantal unieke algoritmen bestaat, het verschil tussen deze documenten te determineren. Het resultaat zal een automatische document-separatie zijn.

Waar komt deze technologie vandaan?

EasyData werkt al meer dan 10 jaar samen met Carp Technologies, dat gespecialiseerd is in taaltechnologie. Overeenkomsten zijn makkelijk te vinden, we werken immers beide met herkenning, alleen op een ander vlak. Nu hebben wij besloten gezamenlijk de herkenningstechnologieën te bundelen in een Semantiek-module. Net als de andere EasySeparate modules kenmerkt deze oplossing zich door een goede kwaliteit/prijs-verhouding.

Moeten alle documenten handmatig geclassificeerd worden?

In bestaande systemen zal dit niet nodig zijn: de classificatie is al bekend aan de hand van bestaande classificatie. Zo zouden offertes, aanmaningen, klant reacties, etc. in het bestaande systeem deze kenmerken al kunnen hebben. Als we (bestaande) document classificatie aan de Semantiek-module kunnen aanbieden, zal training voor een groot deel automatisch verlopen. Met deze aanpak zal de klant snel resultaat hebben van de geleverde inspanningen.

Documentair StructuurPlan (DSP)

Dit is de noemer waaronder deze module valt en biedt organisaties een handvat aan van aanwezige informatiestromen. In een DSP worden relaties gelegd tussen documentsoorten, organisatiestructuur en de juiste werkprocessen. Zo kan een document direct bij de juiste afdeling met eventueel aangevulde gegevens terecht komen. Met een DSP wordt een werkwijze voor de combinatie medewerkers en documenten in de organisatie neergelegd.

Combinatie technologie

Wij zijn al jaren in staat bijvoorbeeld facturen te herkennen door middel van FlexiCapture technologie. Onafhankelijk van de plaats waar bedragen en order-referenties voorkomen kunnen wij deze gegevens extraheren. We blijven deze techniek samen met de al bestaande intelligentie in EasySeparate gebruiken. Er is nu een techniek bijgekomen om documenten ook op hun inhoud te kunnen herkennen of classificeren. Deze bundeling van verschillende disciplines levert nu een breder portfolio van herkenningstechnieken op.