

Op de JCP-borrel tijdens JavaOne liep ik Steve Rondel tegen het lijf. Hij bleek spec lead voor JSR 113, de Java Speech API. Een interview over een wellicht minder bekende, maar niet minder interessante Java-technologie die de komende jaren een grote verbreiding gaat kennen.

Spraaktechnologie naar JME

Interview met Steve Rondel, spec lead JSR 113

Java-spraaktechnologie is er al sinds 1998. De eerste Java Speech API had echter een aantal nadelen. De tweede Java Speech API nadert voltooiing en biedt veel meer praktische toepassingsmogelijkheden. De eerste versie (1998) was gericht op SE, terwijl versie 2.0 zich richt op Java ME. Bij beide versies gaat het om spraakherkenning en spraaksynthese.

Rondel: 'In de tijd van de eerste API waren we geïnteresseerd in desktop deployments. Toen die kleine apparaten op de markt verschenen begonnen we ons af te vragen of sommige van die kleine apparaten genoeg rekenkracht zouden hebben om een trede in de IT-voedselketen af te dalen. Het begon er al snel op te lijken dat mobiele telefoons het primaire doel voor spraaktechnologie zou worden. De microfoons en luidsprekers waren al voldoende geminiaturiseerd en je kon erop gokken dat de processoren genoeg snelheid zouden winnen en dat de draadloze bandbreedte zou gaan groeien.

Esoterisch

Mobiele telefoons hadden toen nog vaak heel esoterische besturingssystemen, ze werkten meestal met eigen gesloten (proprietary) besturingssystemen van de fabrikanten. Niet lang daarna kwam Symbian op nadat Ericson en Nokia samen waren gaan werken, Palm begon, Windows CE wilde meedoen maar kon niet echt klein

genoeg worden en er was nog ene reeks kleinere systemen.

Er waren nog steeds zoveel besturingssystemen en processoren dat er te veel tijd en geld in aanpassingen aan die systemen zou gaan zitten, ook omdat mobiele telefoons toen modelcycli begonnen te kennen van een half jaar. Om stabiliteit in die markt te brengen begon Conversay het idee te promoten om tot een standaard te komen, we vormden een expert groep bestaande uit de belangrijkste Java adapters in de wereld, IBM, Nokia, en Motorola en Texas Instruments en Intel.

Mainstream

De laatste twee waren meer geïnteresseerd in de componentenkant van de mobiele telefoons en Java had zich al gevestigd als een mainstream taal voor applicaties op serverniveau. J2EE servers waren een heel eind en de telefonie providers keken naar mobiele telefoons als een mogelijkheid om aan software en data geld te verdienen, omdat de prijzen voor telefoongesprekken steeds verder begonnen te eroderen.

Dat was geboortegeschiedenis van de standaard groep die we in 2001 opgestart zijn. We vergaderden iedere week gedurende zes jaar met de top spraak wetenschappers in die bedrijven. We zaten met grote uitdagingen. Allereerst hadden de apparaten maar weinig geheugen. Spraak technologie op servers gaf heel matige resultaten, vooral voor mobiele telefoons omdat er

Dré de Man

Tekst en fotografie

vrij veel ruis met het spraaksignaal meege-stuurd werd. De server moest aan de andere kant uit die mix van ruis en gesproken spraak proberen te herkennen en dat werkte niet echt goed.

Wanneer we de *speechengine* in het apparaat onder konden brengen, zouden we in ieder geval de transmissieruis kwijt zijn. Spraakherkenning is bovendien veel beter in staat om te gaan met ruis in de omgeving wanneer de ruisopheffingsalgoritmen op de telefoon geïnstalleerd zouden zijn. Daarnaast zouden we dan applicaties lokaal kunnen laten werken, zodat je niet on line hoefde te zijn om ze te gebruiken.

Zes jaar

Veel voordelen dus, maar het heeft toch nog zes jaar geduurd voor we zover waren. Vorige zomer hebben we een public review van de eerste standaard gepubliceerd en in de herfst hadden we goedkeuring van alle grote Java partijen en we zijn nu een maand van het moment dat de executive committee het uiteindelijke groen licht zal geven. Onze spraakherkenningstechnologie samen met de informatie over hoe die te gebruiken wordt op JavaOne

aan ontwikkelaars uitgereikt, samen met een technologie compatibiliteitskit die een test doet om te kijken of een applicatie voldoet aan de standaarden. JSR 113 zal waarschijnlijk deel gaan kunnen uitmaken van de *wireless toolkit* voor de wireless device makers. Veel mobiele providers zijn er zeer in geïnteresseerd ontwikkelaars te helpen applicaties te ontwikkelen die gebruik maken van spraakherkenning.

Wartoe is de software zoals u die bijvoorbeeld aanbiedt in staat? Hoeveel woorden worden werkelijk herkend bijvoorbeeld
Rondel: 'Wij hebben een technologie die 30.000 woorden herkent. Ons instapniveau spraakherkenning kent 4000 woorden, wat al een veel groter vocabulaire is dan de meeste applicaties ooit zullen gebruiken.' (En veel groter dan dat van de MTV-generatie, DdM)

Er zijn meerdere trends die allemaal wijzen in de richting van een groei van toepassingen van deze technologie. Meer dan vijftig landen kennen beperkingen op het niet handsfree bellen in de auto. Daarnaast zullen we ook spoedig een SMS-verbod voor automobilisten zien. In de enterprise

markt is de toepassing van mobiele telefoons voor computing beperkt door het kleine scherm en het kleine toetsenbord. Het wordt wel gebruikt, maar je moet toch informatie intoetsen. Wanneer je gebruik maakt van spraakherkenning en speechrespons kun je het apparaat in een houder dragen en een headset gebruiken. Je zegt dan iets zeggen in de trant van: 'start inventaris' en het apparaat zal dan antwoorden: 'wat is het onderdeelnummer?', waarop jij zegt 'w23456786798' om vervolgens de vraag naar het aantal producten op het schap te beantwoorden enzovoorts.'

Deze *handsfree eyesfree computerinteractions* zoals wij ze noemen zijn voor heel geschikt voor allerlei groepen: piloten, vrachtwagenchauffeurs, mensen die vliegtuigonderhoud doen, doctoren, verpleegsters, noem maar op. Het maakt het mogelijk taken uit te voeren waarbij je je ogen en handen voortduren nodig hebt terwijl je toch je computer gebruikt. Met een agiele conversatie met de computer is van alles mogelijk. Hetzelfde geldt ook voor productie- of reparatiesituaties; de



Steve Rondel: 'Het meeste werk zit in de wiskundige vergelijkingen die een foneem herleiden tot een letter'



*Compu'Train verzorgt ruim
500 ICT-cursussen op
8 locaties in Nederland.
Jaarlijks begroeten
wij ruim 20.000 cursisten.*

Cursussen

- JavaScript Basis
- JavaScript: Browser-servercommunicatie met AJAX
- Java SE Fundamentals
- Oracle 10g: Java Programming



Compu'Train

Utrecht • Amsterdam • Arnhem • Den Haag
Drachten • Eindhoven • Sittard • Venlo

www.computrain.nl • 0800-2667887


IPROFS **KENNIS**
K JAVATECHNOLOGIE
WALITEIT **T**
NIEUW WERK **PASSIE EN WARMTE**
TEAMGEEST **PROTS**

Steeds vaker valt je op dat er eigenlijk niemand in je directe omgeving is waarvan jij nou écht nieuwe dingen leert. Sterker nog, je bent het grootste deel van de dag bezig om je collega's uit te leggen hoe zij hun vak beter kunnen uitoefenen.

Dat kan toch anders!?

Om je ambitie waar te maken heb je collega's nodig die je scherp houden en open staan om hun kennis en best practices met jou te delen. Samenwerken met dezelfde passie. De Java/J2EE technologie. En bovendien gezien worden als specialist bij opdrachtgevers en in de community. Over ambitie gesproken...

Durf jij het aan om de ICT branche een betere reputatie te geven?

Kom dan snel kennismaken en neem contact op met Ron Brinckman,
Telefoonnummer: 023 – 547 6369
E-mail: rbrinckman@iprofs.nl
Internet: www.IPROFS.nl

Mobiele providers zijn zeer geïnteresseerd in applicaties die gebruik maken van spraakherkenning

computer zegt dan, schroef dit in elkaar en let erop dat je dit lichtje rechts plaats enzovoorts. Je hebt je assistent die je als het ware over je schouder meeprepraat. Er is daar een grote productiviteitswinst te boeken. In bedrijven zullen de potentiële ROI dit soort applicaties aan een snelle groei helpen. Het maakt het mogelijk om zonder papier en toetsenborden te werken, data sneller in te voeren.'

Rondel is echt enthousiast over de mogelijkheden die er voor programmeur en bedrijven zijn op het gebied van spraaktechnologie:

'Applicaties zijn er nog nauwelijks. Het is echt leuk voor een programmeur om je handen op een kennisgebied te leggen, die kennis om te zetten in een conversatieachtige manier en een systeem te bouwen gebaseerd op die kennis. Scholing en educatie is ook een belangrijk toepassingsgebied. Ik praat regelmatig met bedrijven en overheidsinstellingen in Afrikaanse landen. Wanneer ze me in vertrouwen nemen hoor ik vaak: "de meeste van mijn mensen zijn analfabeet". Hoe geweldig zou het niet zijn om een mobiele telefoon te hebben met programma's om te leren lezen en schrijven? In de ontwikkelingslanden is analfabetisme een enorm probleem. Je isoleert een kennisgebied, neem optellen of delen, Frans, aardrijkskunde, wat dan ook en schrijft ene programmaatje voor de mobiele telefoon en miljoenen mensen hebben er toegang toe. Ik vind het prachtig om een de ontwikkelaargemeenschap de mogelijkheid te geven zo kennis ter beschikking te stellen. De mogelijkheden zijn echt heel groot, vergeet niet: een mobiele telefoon van nu is sneller dan een 286.'

Wanneer een ontwikkelaar een applicatie wil schrijven die gebruik maakt van spraaktechnologie, wat moet hij dan doen? Het eerste zou zijn JSR 113 te bekijken, we hebben daar een link naar de Conversay

developer network. We zullen ook de JSR 113 referentie implementaties downloadbaar maken, ze kunnen de toolkit en de core speech technologie van de evaluatieversie gebruiken en ze kunnen aan het werk gaan. Het is echt een moment waarop er heel veel kansen zijn.

Is het werk voor Conversay nu voor het grootste deel af?

Rondel: 'Er is nog veel te doen, we werken nu met 21 talen (waaronder Nederlands sinds 7 maanden, DdM) en nu werken we aan Turks en Hongaars. We passen voortdurend de engines aan. Ze zijn sprekeronafhankelijk, dus er is geen training. Er zit een spreker-adaptief element in de engine, dus hij kan zich meteen aanpassen aan moeilijke dialecten.

Hoe zit het met dicteren?

Rondel: 'We hebben een redelijk goed werkend programma. Toch zal de copmputer nooit slim genoeg zijn om echt et begrijpen wat we zegen, dat heeft veel van doen met het herkennen van wat iemand zegt. Dicteren binnen een bepaalde context gaat heel goed. Maar bij dicteren waarbij het onderwerp volkomen vrij staat, waarin alles aan de orde kan komen, daar beginnen de problemen. Maar wanneer je alleen wil weten wat je moet zeggen wanneer je op reis bent bijvoorbeeld, dat is heel goed te doen. Het zal allemaal groeien, ook dankzij het web.'

De techniek is gebaseerd op het feit dat woorden voor het gehoor zijn opgebouwd uit klankbouwstenen. Deze fonemen hebben ieder een eigen herkenbare frequentie-karakteristiek.

Rondel: 'Iedere taal heeft zij eigen foneem dictionaire. Onafhankelijk van de taal kan onze stem maar een beperkt aantal klanken voortbrengen. Wanneer we een taal bekijken, gebruiken we opnamen van kinderen en volwassenen, mannen en vrouwen en laten die onderzoeken door com-

puter linguïsten die van een bepaalde taal de fonetische compositie kennen en we duiken erin. Om een taal fonetisch in kaart te brengen duurt van drie tot zeven maanden, afhankelijk van de taal, de ene taal is complexer dan de andere. Maar dan komt het echte probleem: wanneer je iets zegt in je taal en wij herkennen de vijf kenmerkende fonemen van het Hongaars, wat zijn de letters die die fonemen voortstellen? Het meeste werk zit in de wiskundige vergelijkingen die een foneem herleiden tot een letter. Dat is een erg fuzzy vergelijking, waarbij het niet zoals meestal gaat om binaire zaken als ja/nee, goed/fout, aan/uit. Het nemen van beslissingen over de mapping van fonemen tot letters gaat gebaseerd op andere fonemen in de buurt van het foneem in kwestie. Zo ken je in het Engels there, their en they're, bij toontalen als het Chinees heb je woorden die op basis van de intonatie een geheel andere betekenis krijgen et cetera. Meestal zijn het tussen de vijftig en 75 klanken die een taal definiëren.

Daarnaast moet je echter ook de tools hebben. Je kunt al die kennis over de talen wel in kaart brengen, maar je moet hem ook kunnen gebruiken. Een systeem moet over de kennis beschikken die kan voorspellen wat de volgende vraag in een gesprek zal zijn bijvoorbeeld. Het gaat om het ontwerp van een expertsysteem. Je vertaalt kennis in een natuurlijk taal dialoog en je combineert het met spraakherkenning en -synthese. Daarna breng je het naar de Java community en zoekt naar toepassingsmogelijkheden voor bijna twee miljard mobiele Java-telefoons.

Meer info:

<http://jcp.org/en/jsr/detail?id=113>