

Contentsyndicatie met RSS en Atom

Standaards nog volop in ontwikkeling

Contentsyndicatie is het beschikbaar stellen van (een deel van) de content van een website aan andere services. Hierbij moet vooral worden gedacht aan het tonen van headlines van nieuwsberichten op andere websites en tools die nieuwsberichten of weblogposts aggregeren. De beschikbaar gestelde content wordt een feed genoemd. Vaak stelt een website meerdere feeds beschikbaar, bijvoorbeeld voor verschillende nieuws categorieën. In dit artikel worden twee standaards, en toepassingen daarvan, tegen het licht gehouden.

Op veel websites is tegenwoordig een oranje XML- of RSS-icoon te vinden. Dit icoon is een link naar een URL waar een RSS document staat dat bij de website of een bepaald deel van de website hoort. RSS is een XML-vocabulary waarmee informatie over (een deel) van de content van een website op een standaardmanier kan worden beschreven. Op basis van de gegevens in het RSS-document kan (een link naar) de betreffende content eenvoudig worden opgenomen in andere websites en beschikbaar worden gesteld aan andere services. Dit wordt contentsyndicatie genoemd en hiermee kan het aantal hits (bezoeken) op een website sterk verhoogd worden.

RSS

RSS is een standaard voor contentsyndicatie. Over RSS bestaat nogal wat verwarring. De verwarring begint al bij het drieletterig acroniem 'RSS'. Oorspronkelijk staat dit voor 'RDF Site Summary' (1.200.000 hits op Google). De eerste versie was namelijk gebaseerd op RDF (Resource Description Framework). Later is de betekenis 'Rich Site Summary' in gebruik genomen (163.000 hits op Google), omdat men zich niet helemaal (meer) aan de RDF specificatie had gehouden. De meest populaire betekenis is echter 'Really Simple Syndication' (3.840.000 hits op Google).

Bij de verschillende versies van RSS slaat de verwarring helemaal toe. De versie nummers volgen elkaar namelijk niet chronologisch op. Dit is voornamelijk te wijten aan het feit dat ver-

schillende groepen, met verschillende inzichten, zich de RSS-'standaard' hebben toegeëigend. De eerste twee versies 0.9 en 0.91 zijn door Netscape ontwikkeld. De eerste was op RDF-gebaseerd, de tweede niet meer. Daarna is Dave Winer van Userland doorgeshaan met de simpele variant van RSS (niet op RDF gebaseerd) en heeft eerst versie 0.92 en daarna 2.0 ontwikkeld. Een andere groep is teruggekeerd naar de RDF-basis. Dit heeft geresulteerd in versie 1.0 die iets eerder dan versie 0.92 uitgebracht is. Dit geeft een aardig inzicht in het eigenzinnige 'wereldje' van standaarden en frameworks. Gelukkig hoeven we ons in de praktijk eigenlijk alleen nog maar te richten op versie 2.0 van RSS, de 'simpele versie', waardoor de betekenis 'Really Simple Syndication' inderdaad het meest van toepassing is.

Atom

Met die laatste (geruststellende) vaststelling zijn we er echter nog niet. Atom is een nieuwe standaard voor contentsyndicatie die sterk in opkomst is. De ontwikkeling van deze standaard is in 2003 in gang gezet door Tim Bray (één van de grondleggers van XML) en Sam Ruby (één van de belangrijkste mensen binnen Apache). Op de weblog van Tim Bray worden de redenen uiteengezet voor het ontwikkelen van (alweer) een nieuwe standaard: <http://www.tbray.org/ongoing/When/200x/2003/06/23/SamsPie>. Hier is te lezen dat de schijnbaar erg eigenzinnige, Dave Winer niet wilde meewerken aan verbeteringen aan RSS (2.0). Atom heeft inmiddels al veel ondersteuning gekregen. Onder anderen van Six Apart, die de veelgebruikte webloggingsoftware Movable Type op de markt brengt. In vergelijking met RSS is Atom completer, duidelijker en een stuk beter gespecificeerd. Atom zal worden ondergebracht bij IETF en daarmee een officiële open standaard worden. Met Atom zal wel degelijk rekening gehouden moeten worden. Het laatste nieuws is overigens dat Dave Winer toenadering heeft gezocht tot Atom.

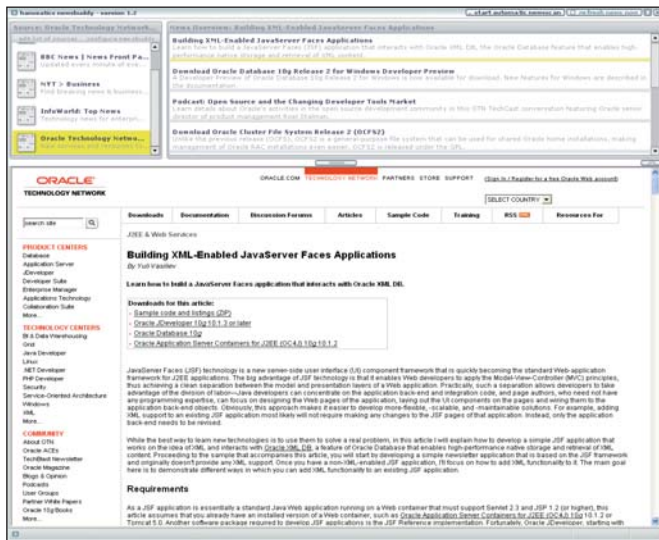
Toepassingen

RSS en Atom zijn op zich breed toepasbaar. Beide standaarden kunnen worden gebruikt om informatie óver content of de

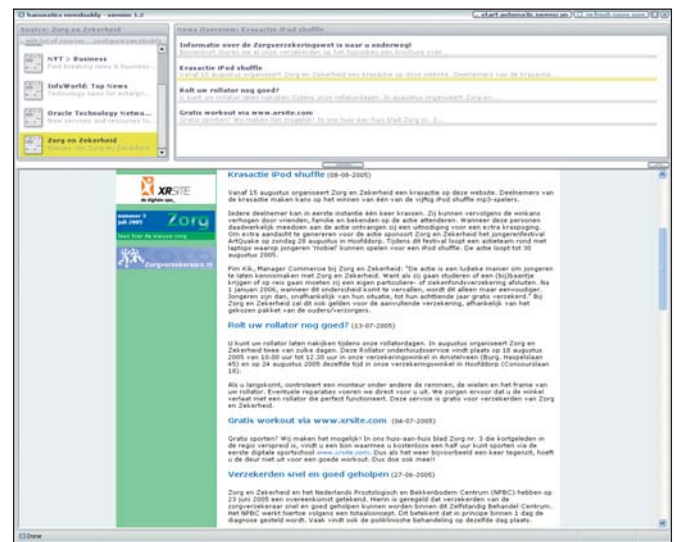
content zélf beschikbaar te stellen. De content kan variëren van nieuwsberichten of weblog-posts tot normale webpagina's. In de praktijk betreft de toepassing echter vooral het beschikbaar stellen van headlines van nieuwsberichten of titels van weblog posts. Vaak wordt er ook een samenvatting beschikbaar opgenomen in de feed. In het algemeen kan gesteld worden dat het gaat om website-content die regelmatig vernieuwd wordt. Het zou dus ook kunnen gaan om nieuwe boeken, DVD's en CD's die gepubliceerd worden, boek- of filmreviews, aanbiedingen, achtergrondartikelen, et cetera. De RSS- of Atom-feed kan gebruikt worden om de content op een andere website op te nemen. In de regel worden dan de headlines, met eventueel de samenvatting, getoond met een link naar de oorspronkelijke website voor het lezen van het complete artikel. Het is meestal immers de bedoeling dat het aantal hits (bezoeken) op de oorspronkelijke website vergroot wordt. Maar ook de website waarop de headlines opgenomen worden kan het aantal hits verhogen doordat het een soort portaal-functie gaat vervullen.

RSS readers

Als Internetgebruiker kun je ook zelf beschikbare feeds gebruiken. Er bestaan namelijk een heleboel tools, *RSS readers*, die als *aggregator* van nieuwsberichten of weblog-posts gebruikt kunnen worden. Met een dergelijke tool kunnen bijvoorbeeld headlines van verschillende nieuwswebsites snel en eenvoudig bekeken worden. Ook geven de meeste tools automatisch een alert wanneer er een nieuw bericht verschijnt op één van de geselecteerde websites (feeds). Het registreren van een feed gaat heel eenvoudig door de URL van de RSS- of Atom-feed in te voeren. De tools verschillen erg in functionaliteit en stabiliteit. Meestal zijn het tools die op de client draaien en op Windows- of Java-gebaseerd zijn. Veelal hebben deze een inge-



Afbeelding 1.



Afbeelding 2.

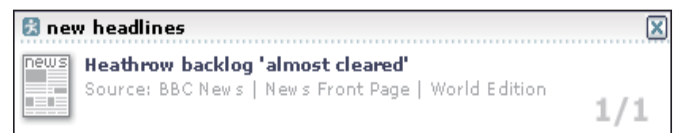
bouwde browser voor het tonen van de webpagina waarnaar verwezen wordt. Ook zijn er webgebaseerde (server-)tools. Tenslotte zijn er ook screensavers die headlines van een bepaalde feed kunnen tonen. In afbeelding 1 en 2 zie je een RSS-feed van OTN en de feed die later in dit artikel gebruikt wordt als voorbeeld met nieuws van Zorg en Zekerheid in de RSS-reader Newsbuddy (niet aan te raden). De RSS-reader in afbeelding 3 toont de pagina waarnaar verwezen wordt in een ingebouwde browser. Er wordt ook een alert getoond wanneer er een nieuwe headline beschikbaar is:

Live Bookmarks in Firefox

Firefox ondersteunt RSS en Atom feeds op een hele handige manier via *Live Bookmarks*. Door binnen de <head> tag van een HTML-pagina <link> elementen op te nemen zoals in onderstaand voorbeeld herkent Firefox dat er RSS- of Atom-feeds beschikbaar zijn.

```
<link rel="alternate" type="application/rss+xml" title="RSS 2.0 Nieuws van Zorg en Zekerheid" href="http://www.zorgenzekerheid.nl/ZENZ_rss2.xml"/>
<link rel="alternate" type="application/atom+xml" title="Atom 0.3 Nieuws van Zorg en Zekerheid" href="http://www.zorgenzekerheid.nl/ZENZ_atom.xml"/>
```

Firefox zal dan rechtsonder in de browser een icon tonen. Door te klikken op dit icon kan één van de aangeboden feeds



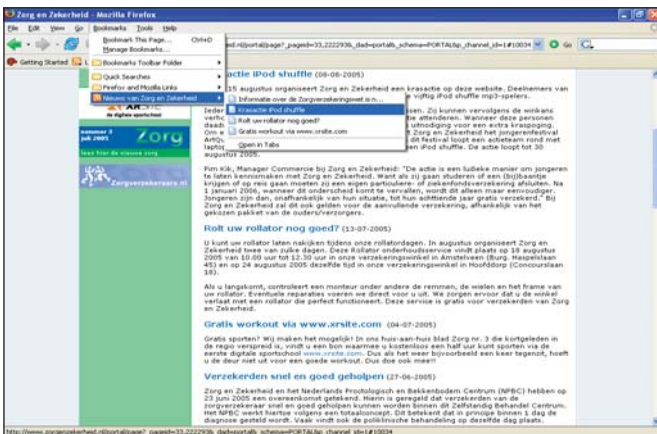
Afbeelding 3.



Afbeelding 4.



Afbeelding 5.



Afbeelding 6.

geselecteerd worden (afbeelding 4 en 5). De feed kan vervolgens als live bookmarks opgenomen worden. De entry's in de feed worden dan als bookmarks getoond.

RSS search-engines

Er zijn ook search-engines die alleen zoeken binnen RSS en Atom feeds zoals Bloglines (www.bloglines.com) en Feedster (www.feedster.com). Om gevonden te worden kan het dus belangrijk zijn om content niet alleen op de website te publiceren maar ook via RSS- en/of Atom-feeds.

Podcasting

Een andere toepassing van RSS is *podcasting*. Met podcasting worden via RSS-audiofiles gedistribueerd (in plaats van verwijzingen naar webpagina's).

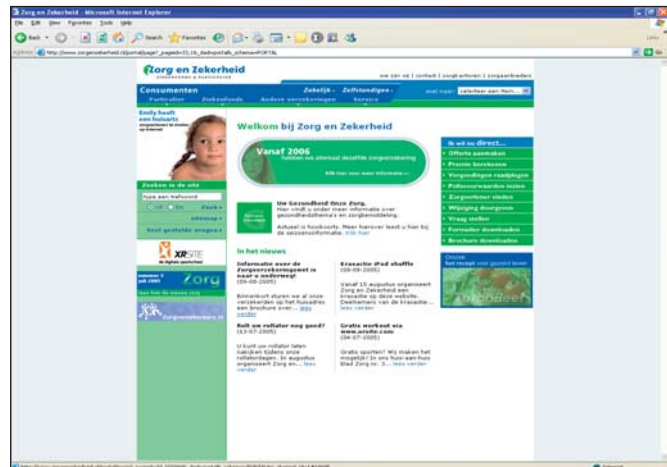
Oracle en RSS

Op www.oracle.com zijn (nog) geen RSS-feeds te vinden, maar op OTN wemelt het van de XML-icons. Op <http://www.oracle.com/technology/syndication/index.html> wordt een overzicht gegeven van de beschikbare RSS-feeds. Er worden (nog) geen

Atom-feeds beschikbaar gesteld door Oracle. Op OTN is een ook RSS reader als extensie op JDeveloper te downloaden. Het Oracle-platform en J2EE bieden prima ondersteuning voor het ontwikkelen van RSS- en Atom-toepassingen.

Ontwikkelen van toepassingen

Het ontwikkelen van contentsyndicatie-toepassingen met RSS en Atom kan betrekking hebben op het beschikbaar stellen van een feed of het lezen en presenteren van een feed. Beiden worden hieronder behandeld. Eerst kijken we naar een voorbeeld van een RSS 2.0 en een Atom-document, met daarin de headlines van de nieuwsartikelen op de Zorg en Zekerheid website.



Afbeelding 7.

RSS 2.0

```
<?xml version="1.0" encoding="UTF-8"?>
<rss version="2.0">
  <channel>
    <title>Zorg en Zekerheid</title>
    <description>Nieuws van Zorg en Zekerheid</description>
    <link>http://www.zorgenzekerheid.nl</link>
  </channel>
  <item>
    <title>Informatie over de Zorgverzekeringswet is naar u onderweg!</title>
    <description>Binnenkort sturen we al onze verzekeren op het huisadres een brochure over...</description>
    <link>http://www.zorgenzekerheid.nl/portal/page?_pageid=33,222293&_dad=portal&_schema=PORTAL&p_channel_id=1#10035</link>
    <pubDate>Tue, 09 Aug 2005 00:00:01 GMT</pubDate>
  </item>
  <item>
    <title>Krasactie iPod shuffle</title>
    <description>Vanaf 15 augustus organiseert Zorg en Zekerheid een krasactie op deze website. Deelnemers van de krasactie...</description>
    <link>http://www.zorgenzekerheid.nl/portal/page?_pageid=33,222293&_dad=portal&_schema=PORTAL&p_channel_id=1#10034</link>
  </item>
</rss>
```

```

        <pubDate>Mon, 08 Aug 2005 00:00:01 GMT</pubDate>
    </item>
    <item>
        <title>Rolt uw rollator nog goed?</title>
        <description>U kunt uw rollator laten nakijken tij-
dems onze rollatordagen. In augustus organiseert Zorg en...</descripti-
on>
        <link>http://www.zorgzekerheid.nl/portal/page?_
pageid=33,222293&_dad=portal&_schema=PORTAL&p_channel_
id=1#10032</link>
        <pubDate>Wed, 13 Jul 2005 00:00:01 GMT</pubDate>
    </item>
    <item>
        <title>Gratis workout via www.xrsite.com</title>
        <description>Gratis sporten? Wij maken het moge-
lijk! In ons huis-aan-huis blad Zorg nr. 3...</description>
        <link>http://www.zorgzekerheid.nl/portal/page?_
pageid=33,222293&_dad=portal&_schema=PORTAL&p_channel_
id=1#10031</link>
        <pubDate>Mon, 04 Jul 2005 00:00:01 GMT</pubDate>
    </item>
</channel>
</rss>

```

In het RSS 2.0-document is informatie opgenomen over de feed zelf, channel in RSS-terminologie: titel, beschrijving en link. Per artikel, item in RSS-terminologie, is een titel, samenvatting, link en publicatiedatum opgenomen.

Atom

```

<?xml version="1.0" encoding="UTF-8"?>
<feed xmlns="http://www.w3.org/2005/Atom">
    <title>Zorg en Zekerheid</title>
    <subtitle>Nieuws van Zorg en Zekerheid</subtitle>
    <link href="http://www.zorgzekerheid.nl"/>
    <link rel="self" href="http://www.zorgzekerheid.nl/ZENZ_
atom.xml"/>
    <updated>2005-08-09T00:00:01Z</updated>
    <id>http://www.zorgzekerheid.nl/atom/1</id>
    <author>
        <name>Redactie</name>
    </author>
    <entry>
        <title>Informatie over de Zorgverzekeringwet is
naar u onderweg!</title>
        <summary>Binnenkort sturen we al onze verzekerden
op het huisadres een brochure over...</summary>
        <link rel="alternate" href="http://www.zorgzeker-
heid.nl/portal/page?_pageid=33,222293&_dad=portal&_
schema=PORTAL&p_channel_id=1#10035"/>
        <updated>2005-08-09T00:00:01Z</updated>
        <id>http://www.zorgzekerheid.nl/atom/1/10035</id>
    </entry>
    <entry>
        <title>Krasactie iPod shuffle</title>
        <summary>Vanaf 15 augustus organiseert Zorg en
Zekerheid een krasactie op deze website. Deelnemers van de krasac-
tie...</summary>
        <link rel="alternate" href="http://www.zorgzeker-

```

```

heid.nl/portal/page?_pageid=33,222293&_dad=portal&_
schema=PORTAL&p_channel_id=1#10034"/>
        <updated>2005-08-08T00:00:01Z</updated>
        <id>http://www.zorgzekerheid.nl/atom/1/10034</id>
    </entry>
    <entry>
        <title>Rolt uw rollator nog goed?</title>
        <summary>U kunt uw rollator laten nakijken tijdens
onze rollatordagen. In augustus organiseert Zorg en...</summary>
        <link rel="alternate" href="http://www.zorgzeker-
heid.nl/portal/page?_pageid=33,222293&_dad=portal&_
schema=PORTAL&p_channel_id=1#10032"/>
        <updated>2005-07-13T00:00:01Z</updated>
        <id>http://www.zorgzekerheid.nl/atom/1/10032</id>
    </entry>
    <entry>
        <title>Gratis workout via www.xrsite.com</title>
        <summary>Gratis sporten? Wij maken het mogelijk! In
ons huis-aan-huis blad Zorg nr. 3...</summary>
        <link rel="alternate" href="http://www.zorgzeker-
heid.nl/portal/page?_pageid=33,222293&_dad=portal&_
schema=PORTAL&p_channel_id=1#10031"/>
        <updated>2005-07-04T00:00:01Z</updated>
        <id>http://www.zorgzekerheid.nl/atom/1/10031</id>
    </entry>
</feed>

```

De structuur van het Atom-document verschilt nauwelijks van het RSS 2.0-document. Er is geen allesomvattend <rss> element, maar ook het Atom-document bevat informatie over de feed zelf en dan een lijst van artikelen, *entries* in Atom-terminologie. Verder verschillen de namen van enkele elementen (sub-title in plaats van description binnen de feed en summary in plaats van description binnen een entry) en wordt een link iets anders aangegeven.

De 'echte' verschillen (vetgedrukt in bovenstaand voorbeeld) zijn: de link naar de Atom-feed zelf, auteursinformatie binnen de feed (maar mag ook binnen entry en er kunnen ook meerdere auteurs opgenomen worden), wanneer de feed voor het laatst veranderd is en een unieke identificatie van de feed en de entry's. De unieke identificaties kunnen door een RSS-reader gebruikt worden om te onthouden of een gebruiker een bepaalde feed of entry al bekeken heeft. In RSS 2.0 kan optioneel via het 'guid' element ook een (globale) unieke identificatie aangegeven worden.

De datum- en tijdformaten verschillen ook. RSS 2.0 gebruikt RFC-822 en Atom RFC-3339. Bovenstaande voorbeelden zijn eenvoudige minimale voorbeelden. Zowel RSS 2.0 en Atom bieden meer mogelijkheden waardoor de toepassingen complexer worden. De voorbeelden zijn echter representatief voor wat in de praktijk beschikbaar wordt gesteld door websites en waarmee RSS-readers overweg kunnen.

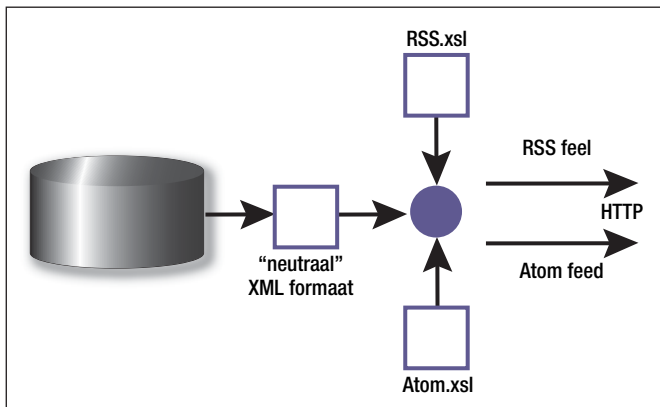
Valideren

Bij het ontwikkelen van RSS- of Atom-feeds is het handig om het geproduceerde XML-document te valideren. Dit kan bij-

voorbeeld op de website <http://www.feedvalidator.org/> van Sam Ruby. Op deze website kun je een URL naar een RSS of Atom-document ingeven. Het document wordt gevalideerd en toont een gedetailleerd overzicht van de eventuele fouten. Als het document valide bevonden wordt, 'mag' je één van onderstaande labels gebruiken.

Beschikbaar stellen van een feed

Om bepaalde content via zowel een RSS 2.0- als een Atom-feed beschikbaar te stellen, kan gebruik worden gemaakt van een 'neutraal' XML-formaat dat alle informatie bevat en met behulp van XSLT-stylesheets (RSS.xml en Atom.xml in afbeelding 8) omgezet kan worden naar de RSS 2.0 of Atom flavor. Het samenstellen van het XML-document zou met behulp van SQL/XML gedaan kunnen worden (zie Optimize 3/2004) en de transformatie kan uitgevoerd worden via de transform-method van XMLType (zie Optimize 4/2004). De feed kan als een servlet beschikbaar worden gesteld of om de zoveel tijd als een statische file gepubliceerd worden.



Afbeelding 8.

Lezen en publiceren van een feed

Voor het lezen van een feed kan gebruik worden gemaakt van HTTPURIType. Dit type maakt gebruik van de UTL_HTTP database package om het resultaat van het aanroepen van een URL over http op te halen. Met de getXML-method van dit type kan de feed in een XMLType ingelezen worden. Vervolgens kan de XML met behulp van een XSLT-stylesheet omgezet worden naar HTML. In de volgende PL/SQL-code wordt dit geïllustreerd:

```

declare
    l_rss_document      xmltype;
    l_xslt_stylesheet  xmltype;
    l_html_document    xmltype;

begin

```

```

    l_rss_document := HTTPURIType(p_url).getXML();

    l_xslt_stylesheet := XDBURIType('/public/CNE/rss2_to_html.xml').
getXML();

    l_html_document := l_rss_document.transform(l_xslt_stylesheet);

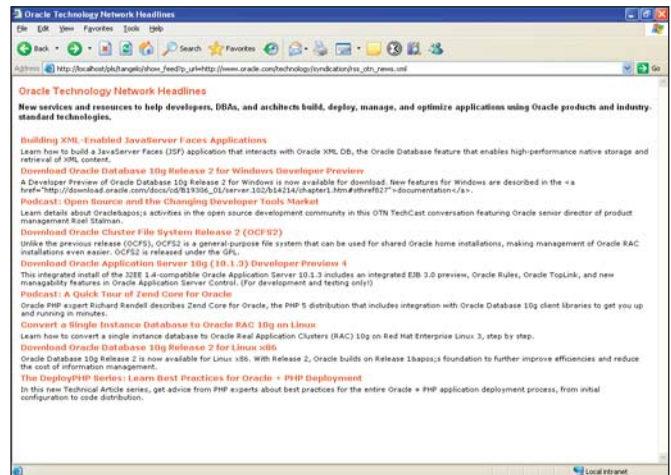
exception

when others
then
    dbms_output.put_line(sqlerrm);

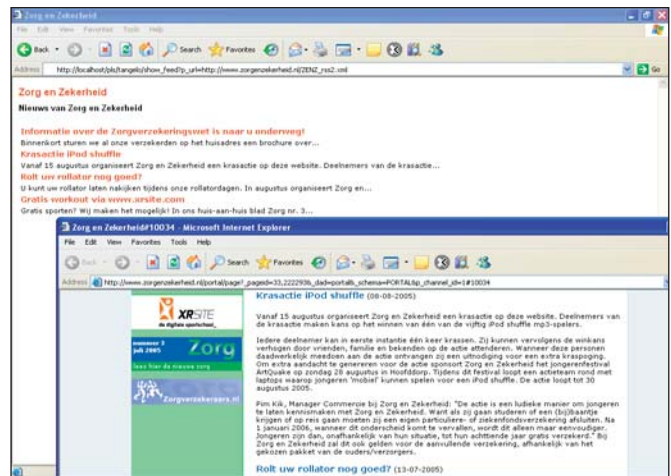
end;
/

```

Bij gebruik in een servlet resulteert dit in de volgende pagina's voor de Oracle Technology Network Headlines feed (http://www.oracle.com/technology/syndication/rss_otn_news.xml) en het nieuws van Zorg en Zekerheid voorbeeld uit dit artikel, zoals te zien in afbeelding 9 en 10.



Afbeelding 9.



Afbeelding 10.

Oracle Portal

Oracle Portal biedt standaard geen ondersteuning voor het beschikbaar stellen van RSS-feeds voor content op een portaal. De Cumquat Basismodules bieden standaardondersteuning voor zowel RSS 2.0 als Atom-feeds voor nieuwsberichten en Document Editor biedt deze feeds voor gepubliceerde documenten. Voor het lezen en publiceren van feeds biedt Oracle Portal met het standaard XML-portlet beperkte ondersteuning. Cumquat biedt als onderdeel van haar standaard Oracle Portal-implementatie componenten om op een eenvoudige manier zowel RSS 2.0 als Atom-feeds te lezen en te publiceren.

Erwin Groenendal is Senior Solutions Architect bij Cumquat Information Technology (www.cumquat.nl) in Zeist. Vragen en opmerkingen over dit artikel kunnen gestuurd worden aan erwin@cumquat.nl.



Call for Speakers

DATABASE SYSTEMS 28-30 MAART 2006

DATABASE SYSTEMS 2006
De dertiende editie van het bekende IT-evenement **Database Systems** vindt plaats van 28 tot en met 30 maart 2006. Dit jaarlijkse evenement is gewijd aan onderwerpen als softwareontwikkeling, systeemintegratie, databasetechnologie, datawarehousing, architectuur en trends. Iedere congresdag zijn er drie parallelle tracks waarin diverse onderwerpen in detail door onafhankelijke specialisten worden besproken.

Voor deze editie wordt een **Call voor Speakers** uitgeschreven. Heeft u interesse in het geven van een lezing of weet u een geschikt onderwerp, doe dan vóór 31 oktober 2005 een voorstel!

Geïnteresseerd? Ga dan snel naar www.databasesystems.nl voor meer informatie en voor het volledige inschrijfformulier.




We doen
in ICT.

Jij hebt
energie.

Sogeti is één van de top-5 ICT-bedrijven van Nederland. Wij staan garant voor interessant werk, met Oracle-opdrachten bij interessante klanten. We profileren ons daarbij met kwaliteit en vakmanschap. Klanten zien dit terug in onze dienstverlening en in de door ons ontwikkelde en tot standaard uitgegroeide methoden zoals DYA®, Regatta®, TMap® en TPI®.

Innovatieve engineering met Oracle

Binnen de divisie Distributed Software Engineering (DSE) werken engineers die gebruik maken van 'leading technology', waaronder Oracle. Onze Oracle-community bestaat uit ruim 70 enthousiaste professionals die resultaatgericht werken en met heel hun hart voor het ontwikkelen van systemen gaan. We maken klantgerichte oplossingen, waarbij we werken met onder andere J2EE, Oracle 8i, 9i, 9iAS, 10g, Oracle DBA-kennis, Oracle Warehousebuilder en Discoverer, JDeveloper en JHeadstart. Ook mobiele oplossingen op basis van het Oracle-platform ontbreken niet.

Gezien de sterke stijging van het aantal opdrachten op het Oracle-vakgebied zoeken wij:

- Ervaren applicatieontwikkelaars
- Ervaren DBA's
- Young Professionals

We zoeken voor deze functies vakmensen met een HBO-werk- en denkniveau en aantoonbare ervaring die klinkende resultaten kunnen behalen door onze klanten te ondersteunen en te adviseren omtrent het gebruik en de mogelijkheden van Oracle. Wij bieden je goede arbeidsvoorwaarden en uitstekende opleidings- en certificeringsmogelijkheden.

Ben je geïnteresseerd? Stuur dan je cv met motivatie naar dse-hrm@sogeti.nl. Heb je nog vragen, dan kun je bellen met onze HR manager Merie Sigmond op 020 660 66 00 + nakiesnummer 7614.

Sogeti Nederland B.V. Postbus 76, 4130 EB Vianen
Tel (020) 660 66 00 Fax (020) 660 67 21 www.sogeti.nl



SOGETI
Realisme in ICT