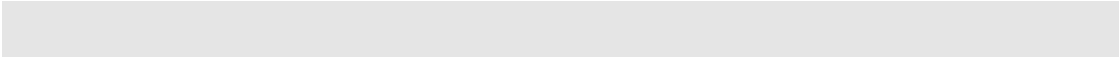


RSS

Laborationskompendium för syndikering på Internet

Detta exemplar tillhör: _____

VT 2005



Laborationskompendium i RSS 2.0

Inledning

Vi lever nu i vad man kallar för informationssamhället. Aldrig har vi konfronterats med så mycket information. Dagligen tar vi del av information från olika källor såsom radio, tidningar, TV, reklam, e-post etc. Internet är dessutom en informationskälla som innehåller en oöverskådlig mängd information. Problemet är ju inte längre att få tillgång till information, utan snarare att kunna hitta något intressant av allt som erbjuds. Hur ska vi på ett enkelt sätt kunna ta del av den information som webben erbjuder? Med hjälp av sökmotorer kan man söka efter information men tyvärr så blandas relevant information med rent skräp. De som ofta använder sig utav Internet har utvecklat sina strategier att söka information på olika sätt. När man väl hittar intressanta webbsidor så gäller det att komma ihåg dessa och regelbundet besöka dessa sidor för att se ifall något nytt har dykt upp. Proceduren att besöka en massa sidor för att se ifall något nytt publicerats kan ta mycket tid i anspråk och det är inte säkert att ny information publicerats. Vore det inte bättre ifall man kunde få dessa nyheter direkt, skickade till sig utan att behöva besöka dessa sidor? Då skulle vi vara uppdaterade utan att aktivt behöva söka och kontrollera om ny information lagts till. Du som är vaken har kanske observerat någon av följande symboler när du varit ute och surfat.



Dessa symboler innebär att den sida du befinner dig på erbjuder dig som användare att ta del av de nyheter som publiceras i form av RSS. Du kan alltså gratis prenumerera på nyheter. När ny information läggs till på webbsidan får du snabbt kännedom om detta i form av en rubrik och en kort beskrivning av innehållet. I anslutning till nyheten finns dessutom länkar till källan så att du kan läsa artikeln eller vad det nu är i sin helhet.

Vad är RSS?

RSS är en teknik som möjliggör enkel spridning av information till flera olika användare. RSS är en akronym och har två utbredda betydelser:

Rich Site Summary: När Netscape (den gamla webbläsaren) skapade RSS 1999, var det i syfte att kortfattat beskriva innehållet på en webbsida. På den tiden sysslade företaget med att bygga nyhetsportaler vars innehåll kunde komma från fler olika källor. Netscape lade ner sitt projekt eftersom det började gå knackigt för dessa nyhetsportaler. Olika grupper och organisationer såg dock potential och möjligheter med denna teknik och fortsatte arbetet med att förenkla och vidareutveckla det som Netscape hade dragit igång. Arbetet har lett fram till en standard som likt HTML finns i olika versioner.

Really Simple Syndication: Syndikering innebär att en webbplats kan sprida och dela med sig av information och nyheter till andra webbplatser. På susning.nu kan man hitta följande definition av syndikering. "Handel med informationsinnehåll som från en central källa infogas i flera oberoende informationsprodukter".

Syftet med RSS är alltså konstruerat för att sprida och offentliggöra information till dem som är intresserade. RSS är, precis som XHTML som ni känner till, en tillämpning av XML. Vad XML är kommer att vi att titta lite närmare på lite senare.

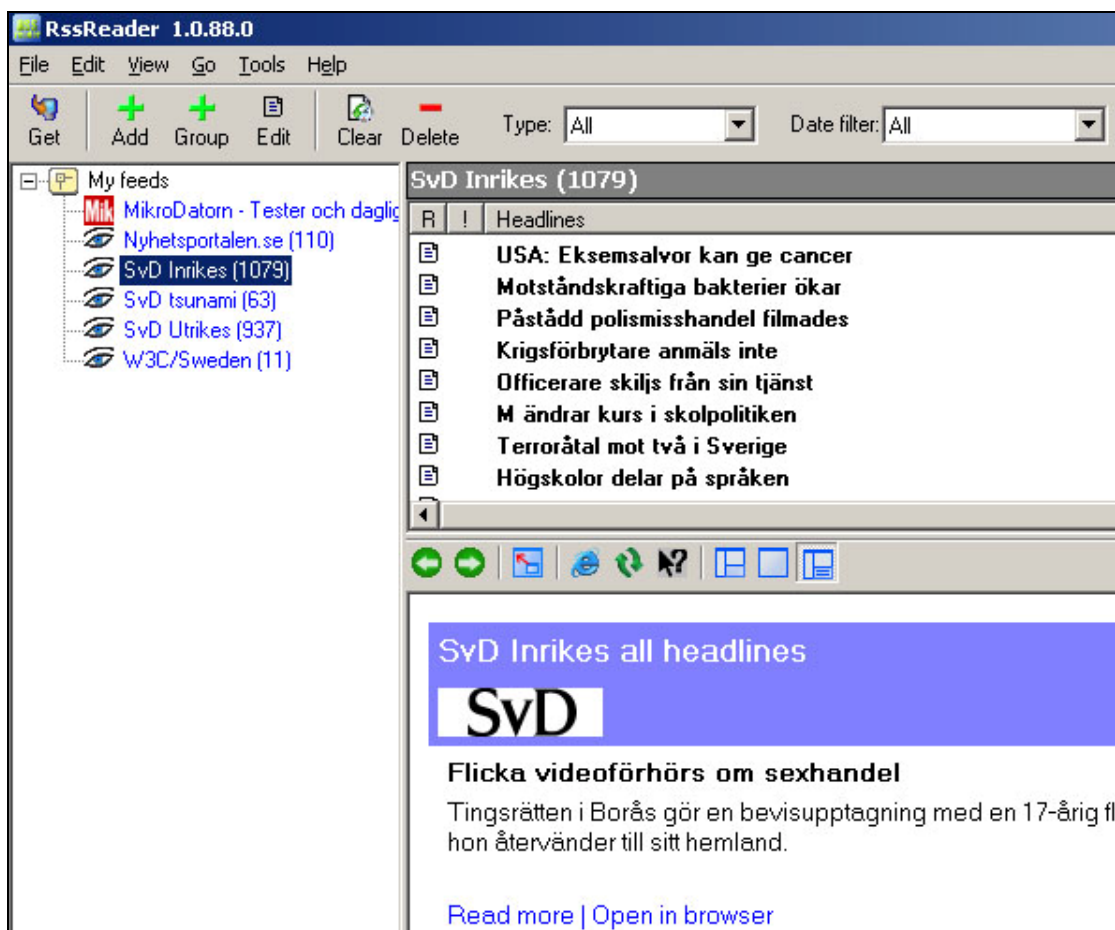
Så här beskriver Computer Swedens ordlista RSS:

RSS - Really Simple Syndication - ("riktigt enkel nyhetsbyrå")

- teknik för automatisk sändning av nyheter och annat webbmateriel från en webbplats till mottagare. En ny och enkel variant av pushteknik. Användaren installerar ett enkelt program, en news reader, som hämtar information enligt användarens inställningar från RSS-anpassade webbplatser. Webbplatserna levererar informationen som RSS feeds, enkla textfiler som inte belastar nätet eller datorn. RSS utvecklades av Netscape och använder ett xml-baserat format.

Varför ska jag använda RSS?

Det viktigaste tillämpningsområdet av RSS är möjligheten att snabbt och enkelt kunna sprida och ta del av de nyheter som publiceras. För de människor som har ett stort behov av att ta del av aktuell information från olika källor kan det vara tidsödande att behöva besöka dessa webbplatser regelbundet för att se ifall några nyheter publicerats. RSS möjliggör för användare att snabbt och enkelt ta del av de nyheter som publiceras. Man kan jämföra det med att prenumerera på nyhetsbrev.

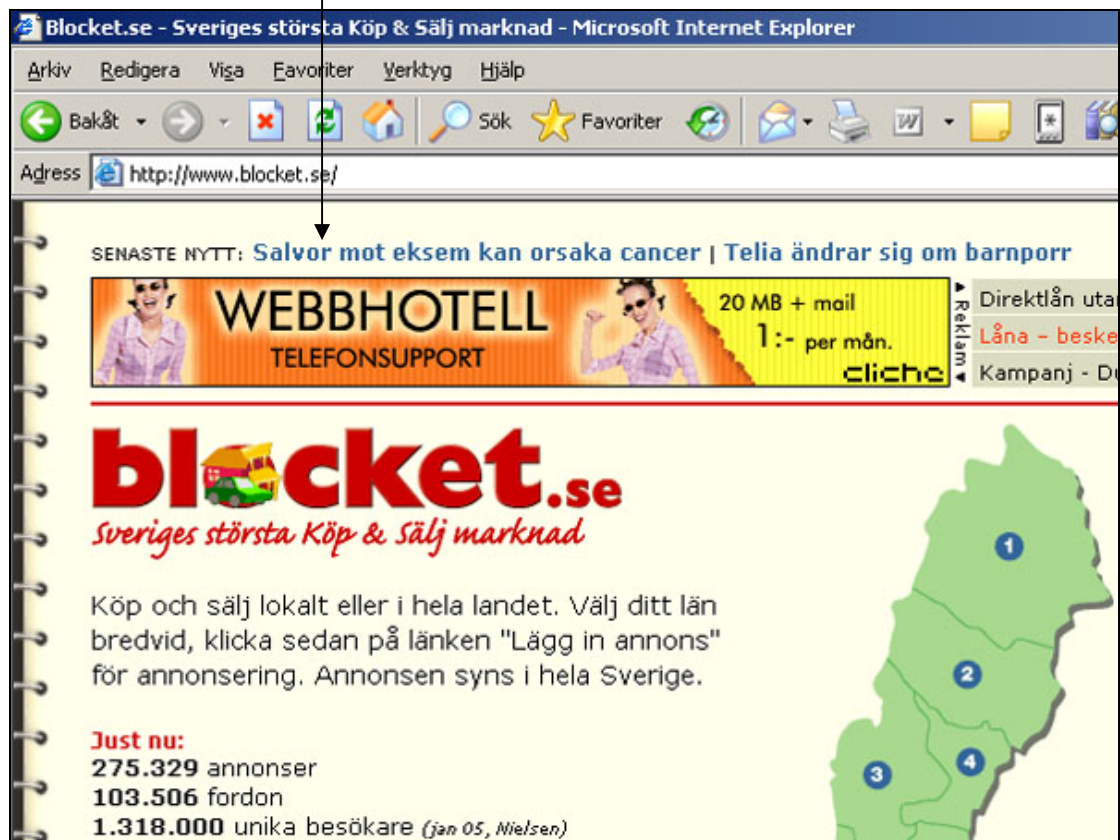


I bilden ovan kan du se att användaren prenumererar på nyheter från olika webbplatser. Uppe till vänster ser du rubriker för aktuella nyheter från SvD inrikes. Strax under ser du rubrik samt en sammanfattning av nyheten samt den länk till SvD där du kan hitta artikeln i sin helhet. Man kan göra inställningar i programmet så att det regelbundet kontrollerar ifall det kommit några nyheter, exempelvis varje timme. Och så fort nyheter läggs till så får man veta detta. Användaren kan därefter utifrån titel och den kortfattade beskrivningen avgöra

ifall nyheten är intressant att läsa i sin helhet eller inte. Programmet ovan heter RssReader och är gratis att använda.

Ett annat användningsområde är att man kan publicera dessa nyhetsrubriker på sina egna webbsidor. Det kan vara ett bra sätt för att locka tillbaka användare till sin webbplats om användarna ser att sidan är uppdaterad och visar aktuella nyheter från någon källa. Detta är en möjlighet som många webbplatser använder sig utav, exempelvis annonstorget Blocket och Internetportalen Spray. Nedan ser du hur Blocket publicerar nyheter som är hämtade från Svenska Dagbladet.

SENASTE NYTT: Nyheterna kommer direkt från Svenska Dagbladet



Hur kan jag använda RSS?

För att "prenumerera" på nyhetstjänster behöver man någon form av program, även kallad RSS-aggregator. Dessa program kontrollerar regelbundet de webbplatser jag valt att bevaka för att se ifall något nytt lagts till. De "nya" nyheterna hämtas därefter hem och presenterar dessa i ett läsvänligt format. Det finns många olika program för detta ändamål. Somliga kostar pengar men det finns också en hel del gratisprogram att ladda ned från nätet. Följande program är bara några exempel på så kallade RSS-aggregatorer som inte kostar några pengar.

- <http://www.rssreader.com> (Windows)
- <http://www.bloglines.com> (webbaserad)

Det finns dessutom så kallade insticksprogram för olika e-postprogram som gör att man kan få dessa nyheter skickade till sin e-post istället.

Uppgift 1

Besök följande länkar och titta på hur så kallade RSS-flöden ser ut.

<http://www.piratbyran.org/piratbyran.rss>

<http://svt.se/svt/road/Classic/shared/xmlexport/rss/newsrss.jsp?d=21336>

<http://www.svd.se/articleexport/rss20/inrikes.xml>

Inte särskilt rolig läsning, eller hur? En massa text i ett nästan oläsbart format, vad ska man med detta till? Om du tittar lite noggrannare så hittar du så kallade taggar som påminner en hel del om HTML-kod som du kanske har sett tidigare. Ett RSS-flöde är helt enkelt nyheter paketerade på ett speciellt sätt i ett bestämt format. Vi kommer att titta lite närmare på det lite senare i labben.

Uppgift 2

Besök Bloglines hemsida på adressen www.bloglines.com

Registrera ett konto (det är gratis) och sök efter intressanta källor som erbjuder information i RSS-format. Lägg till dessa RSS-flöden eller 'Feeds' som de kallas på engelska genom att klicka på fliken 'My Feeds' och välja 'Add'. I det fönster som nu dyker upp anger du länken till de RSS-flöden som du hittat och vill prenumerera på. Bloglines har även en stor förteckning över källor som levererar RSS-flöden. De har även en lista på de mest populära RSS-flödena.



Hur kan jag publicera andras nyheter på mina webbsidor?

Som vi tidigare sett så kan webbsidor hämta aktuella nyheter från olika nyhetskällor. Detta förutsätter att man måste skapa så kallade dynamiska webbsidor. En dynamisk webbsida kan till skillnad från en statisk sida växla utseende oberoende av att man måste skriva om den kod som utgör grunden för hemsidan. En statisk sida ser alltid lika ut och växlar inte utseende såvida vi inte skriver om HTML-koden.

Vi vill ju kunna publicera aktuella nyheter utan att ändra i webbsidan oavbrutet och för att åstadkomma detta så måste man använda sig av något programspråk. Det finns flera möjligheter att skapa dynamiska hemsidor. Man kan skilja mellan två typer av språk. De som körs i webbläsaren, ex JavaScript, och de som körs på webbservern, ex PHP, ASP etc. Vi kommer att använda oss utav PHP vilket är ett serverbaserat programspråk. Det innebär att den webserver där vi har våra webbsidor måste klara av att köra PHP-filer.

I figuren nedan ser du ett exempel på hur PHP-kod kan se ut. Koden är, för ett otränat öga, ungefär lika läsbar som ren grekiska. Men var lugn, ni behöver inte förstå hur PHP-koden fungerar. Som ni ser så finns det också lite HTML-kod i filen. Dessutom finns det två stycken webbadresser, sk URL. Det är den andra adressen som är av intresse. Den hänvisar nämligen till nyheterna.

```
<?
require_once 'http://www.itm.mh.se/~jorsod/magpierss/rss_fetch.inc';

$url = 'http://svt.se/svt/road/Classic/shared/xmlexport/rss/newsrss.jsp?d=21336';
$rss = fetch_rss($url);

echo "Site: ", $rss->channel['title'], "<br>\n";

foreach ($rss->items as $item ) {
    $title = $item[title];
    $url   = $item[link];
    $description = $item[description];
    echo "<a href=$url>$title</a><br>$description<br><br>\n";
}
?>
```

Ovanstående fil kan du hitta på adressen:

http://www.itm.mh.se/~jorsod/nyheter_via_php.html

1. Kopiera texten och klistra in den i en tom textfil. (Använd ex Anteckningar)
2. Spara filen i din egen webbkatalog. (Mappen public_html i ditt portalkonto)
3. Döp filen till nyheter.php. (Filnamnet måste sluta på .php om vi ska kunna köra php-koden)
4. Titta på filen i en webbläsare. Ange adressen till din hemsida, ex portal.miun.se/~ditt_konto/nyheter.php

Vad händer? Jo, inget annat än att den obegripliga koden skrivs ut. Detta beror på att den webserver vi nu använder inte klarar av att köra PHP-filer.

Ange istället adressen http://rainbow.itm.mh.se/~ditt_konto/nyheter.php Nu försvinner PHP-koden och ersätts med de önskvärda nyheterna.

Om du istället vill publicera andra nyheter än de som nu visas så kan du enkelt ändra på det genom att byta ut adressen på raden

```
$url = 'http://svt.se/svt/road/Classic/shared/xmlexport/rss/newsrss.jsp?d=21336';
```

mot adressen till den önskvärda RSS-filen.

Uppgift 3

Du ska nu skapa eller bygga vidare på din webbsida och göra den mer intressant genom att publicera aktuella nyheter på den. I princip kan du kopiera filen ovan och klistra in den på din HTML-fil där du vill placera nyheterna. Kom ihåg att filnamnet måste sluta med *.php* och att du måste titta på filen när den ligger på en webbserver som klarar av PHP, dvs *rainbow*.

Hur är RSS uppbyggt?

RSS är uppbyggt av ett språk som heter XML. XML är en teknik som tagits fram för att strukturera och organisera data/information. XML betyder eXistensible Markup Language och som namnet antyder så är det ett uttänjbart/utbyggbart språk för att märka upp saker. XML är ett så kallat metaspråk, dvs det är inget riktigt språk men kan användas för att definiera nya språk, ex XHTML och RSS.

Vi kan börja med att titta på en enkel tillämpning av XML. I exemplet nedan har man försökt strukturera innehållet i ett meddelande:

```
<meddelande>
  <till>Anders And</till>
  <från>Vilde Vilma</från>
  <rubrik>Filmtajm?</rubrik>
  <innehåll>Hej! Hur mår du? Ska vi gå på bio
    ikväll?
  </innehåll>
</meddelande>
```

Här ovan kan vi se hur man har identifierat de olika delarna i meddelandet. Man har skiljt på avsändare, mottagare, rubrik och själva meddelandet. Detta är en process som är ganska naturlig för oss människor. Vi vill gärna sortera och strukturera information. XML är framtaget för att kunna skapa dessa språk i syfte just att strukturera, beskriva och sprida data. XML:s primära uppgift är alltså att strukturera data.

Ni som tidigare har gjort hemsidor känner säkert igen exemplet ovan tycker kanske att det påminner om språket HTML. Faktum är att den nya versionen av HTML, som fått namnet XHTML, har skapats med XML. XML-applikationer och XML-tillämpningar är vanligt förekommande begrepp som innebär att ett språk definierats med hjälp av XML.

Ett XML-dokument är inget annat än en textfil med ändelsen .xml och som innehåller information uppmärkt med något i XML definierat språk. Ett sådant dokument består normalt av följande delar:

- Rotelement
- Element
- Attribut

Element

Innehållet i ett XML-dokument delas in i så kallade element med hjälp av taggar. Ordet tagg har man lånat från engelskans tag som i det här fallet betyder sätta lapp på, märka upp. Taggar skrivs inom tecknen mindre än < och större än >. För att märka upp något så använder man dessa taggar i början och slutet av det som ska märkas upp. Man kan därför prata om start- respektive sluttaggar. Start- och sluttagg ser lika ut förutom att sluttaggen inleds med ett snedstreck /.

Ex:

```
<meddelande>innehåll</meddelande>
```

Varje XML-dokument har endast ett rotelement. I exemplet ovan så är det meddelande som är rotelementet. I rotelementet kan det sedan förekomma flera element. Rotelementet *meddelande* innehåller följande element:

- till
- från
- rubrik
- innehåll

Attribut

Förutom element så kan ett XML-dokument även innehålla attribut. Ett attribut används för att lägga till information till ett element. Vi kan titta på ett nytt exempel:

```
<person kön="man">  
  <förnamn>Anders</förnamn>  
  <efternamn>And</efternamn>  
  <telefon typ="mobil">555-123 456</telefon>  
  <telefon typ="hem">063-978 645</telefon>  
</person>
```

Här har man lagt till flera attribut. Elementet person har kompletterats med attributet kön som i detta fall har värdet "man". Elementet telefon har ett attribut som heter typ. I exemplet har vi två telefon-element och dessa har olika värden på attributet typ.

(Validering)

Eftersom XML används för att strukturera information så är det väldigt viktigt att man skriver korrekt XML som inte kan misstolkas och är välstrukturerad. (Om vi till exempel väljer en webbsida skriven i XHTML så vill vi ju att den ska se ut på ett speciellt sätt. För att uppnå detta måste vi följa de regler man satt upp, annars kan inte datorn förstå hur den ska visa webbsidan. Utan dessa regler skulle man inte kunna visa sidor på ett vettigt sätt.) Man kan kontrollera att ett XML-dokument uppfyller de regler som satts upp genom att använda så kallade validatorer. En validator är ett program som kontrollerar att dokumentet är korrekt och anger dessutom var och vad som är fel så att man enkelt kan åtgärda sitt dokument. Vi kommer inte att gå djupare än så när det gäller validering.

RSS – En tillämpning av XML

Det finns olika versioner av RSS, men vi kommer att titta närmare på version 2.0.

Rotelementet i denna XML-applikation heter *channel* och underelementen till channel är:

- title
- link
- description
- language
- item

Det viktigaste inslaget i ett RSS-flöde är ju trots allt själva nyheten. Därför tittar vi lite närmare på underelementet *item*.

Exemplet nedan är hämtat från Svenska dagbladets RSS-flöde:

```
<item>
  <link>http://www.svd.se/go.php?sgo/rss/http://www.svd.se/
    dynamiskt/inrikes/did_9336061.asp
  </link>
  <title>Restaurangrånare band personal</title>
  <description>Vid halv åtta i morse rånades en restaurang i
    Stockholm. Sex eller sju människor i personalen bands med
    silvertejp av rånarna som var beväpnade med pistoler och
    baseballträn.
  </description>
  <category>inrikes</category>
  <pubDate>Mon, 14 Mar 2005 08:21:00 N</pubDate>
</item>
```

Så här är alltså en ser en nyhet ut när den definierats i RSS. Varje nyhet är ett *item*. I ett item finns de underliggande elementen: *link*, *title*, *description*, *category* och *pubDate*.

- link = Innehåller länk till fullständig artikel
- title = Artikelns rubrik eller titel
- description = Artikel i kort variant
- category = Kategori
- pubDate = Datum för publicering

Till dessa nyheter tillkommer lite information om själva källan. Dessa utgörs av egna element.

```
<rss version="2.0">
  <channel>
    <title>SvD Inrikes</title>
    <link>http://www.svd.se</link>
    <description>Inrikesnyheter från
      www.svd.se
    </description>
    <language>sv</language>
    <copyright>SvD 2005</copyright>
    <lastBuildDate>Mon, 14 Mar 2005 10:09:23N
    </lastBuildDate>

    <item>
      ... (Här kommer nyheterna, se ovan)
    </item>
  </channel>
</rss>
```

Om du skriver in följande adress i webbläsaren så kan du se de senaste inrikesnyheterna från Svenska Dagbladet:

<http://www.svd.se/articleexport/rss20/inrikes.xml>

Uppgift 4

Gör din egen RSS-fil som innehåller 4-5 nyheter som du vill dela med dig utav. Publicera den i din webbkatalog (public_html) och gör den tillgänglig för allmänheten.

Uppgift 5

Gör om din webbsida så att den publicerar nyheter från en klasskompis RSS-flöde. Det du behöver känna till är adressen till RSS-filen.