

Documenti ISTAT

RSS: uno standard per diffondere informazioni

F. Rizzo, L. Vignola (*)

(*) ISTAT – Direzione Centrale per la Diffusione della cultura e dell'informazione statistica

1. INTRODUZIONE

2. IL PROTOCOLLO RSS

2.1 I VANTAGGI NELL'USO DELL' RSS

2.2 I PROBABILI SCENARI DI RSS

2.3 L'IMPLEMENTAZIONE DEL PROTOCOLLO RSS

2.4 STRUTTURA DI UN DOCUMENTO RSS

2.4.1 Gli elementi del nodo channel

2.4.2 Gli elementi dei nodi item

2.4.3 La generazione dei FEED RSS

2.4.4 I client RSS

3. DUE APPLICAZIONI SVILUPPATE

3.1 I FEED RSS PER LA BANCA DATI CONISTAT

3.2 PROTOTIPO DI UNO STRUMENTO PER SCRIVERE FEED RSS

APPENDICE A

APPENDICE B

APPENDICE C

APPENDICE D

1. Introduzione

RSS, acronimo di Rich Site Summary oppure di Really Simple Syndication o ancora di RDF Site Summary, rappresenta un protocollo, implementato attraverso una semplice grammatica XML (eXtensible Markup Language), per diffondere informazioni sul WEB. Il suo uso iniziale come strumento ideale per diffondere news attraverso dei canali (news syndication), si è evoluto tanto da rappresentare un ottimo meccanismo per condividere informazioni (data sharing).

Migliaia di siti WEB usano oramai RSS come mezzo per diffondere le novità e aumentare la propria visibilità.

Questo documento è stato organizzato logicamente in due parti.

La prima parte introduce il protocollo RSS evidenziando le diverse versioni attualmente esistenti, i vantaggi della sua adozione, le sue possibili applicazioni; la seconda parte espone alcune esperienze realizzate.

Seguono l'appendice A, B e C in cui vengono riportati parti di codici e dettagli tecnici utilizzati per la realizzazione delle varie esperienze presentate, mentre nell'appendice D viene fornito un piccolo glossario e una raccolta di indirizzi di siti web a cui il lettore potrà fare riferimento.

I contributi degli autori al documento sono nel seguente ordine:

capitolo: 1, 2 – Francesco Rizzo

capitolo: 3 – Laura Vignola

Appendice A, D – Francesco Rizzo

Appendice B, C – Laura Vignola

2. Il protocollo RSS¹

Volendo dare una semplice definizione di RSS si potrebbe dire che rappresenta un protocollo per la distribuzione di liste di hyperlink sul WEB. In generale oltre a distribuire hyperlink, RSS aggiunge altre informazioni, dette metadati, che aiutano il fruitore a decidere se percorrere i link.

RSS, come qualunque protocollo, è stato implementato nel tempo in diverse versioni, ciascuna delle quali associa al protocollo una sigla differente. Così RSS può significare RDF (Resource Description Framework) Site Summay nel caso si utilizza la versione 1.0, oppure Really Simple Syndication nel caso si utilizzano le versioni 0.9x o 2.0.

La prima interpretazione considera RSS come un sommario o indice particolarmente ricco delle informazioni presenti su un dato sito WEB. La seconda interpretazione suggerisce che RSS è un “syndication” molto semplice, cioè uno strumento per diffondere informazioni all'interno di un canale, in questo caso specifico Internet. In generale potremmo pensare a RSS come a un modo semplice per propagare “distillati” del proprio sito WEB, veicolando a partire da esso liste di link e brevi estratti di news.

Spesso gli utenti che navigano su Internet visitano periodicamente dei siti WEB per accedere alle novità o agli aggiornamenti su argomenti di loro interesse. Questo lavoro richiede del tempo e in alcuni casi può non produrre alcun risultato in quanto è possibile che non sia stato pubblicato nessun aggiornamento circa l'argomento d'interesse. Per evitare che ciò avvenga molti siti WEB mettono a disposizione degli utenti sommari delle notizie che vengono pubblicate in formato RSS. Tali sommari in letteratura vengono denominati FEED.

L'utente che vuole usufruire di tale servizio dovrà semplicemente installare sul proprio computer un piccolo programma gratuito detto “news aggregator” per importare, ad intervalli regolari, tali FEED visualizzandoli direttamente sulla “scrivania” del proprio computer attraverso finestre simili a quelle utilizzate dai programmi per la “messaggistica immediata”. Chiunque può attingere direttamente alla fonte, creandosi una selezione personale di news e sorgenti informative da consultare.

¹ In seguito i termine protocollo RSS, grammatica RSS, FEDD RSS, formato RSS, documento RSS, file RSS verranno utilizzati indifferentemente, con l'intenzione di indicare lo stesso concetto, così come avviene in letteratura.

2.1 I vantaggi nell'uso dell' RSS

Per analizzare correttamente i vantaggi che l'uso di RSS apporta, è necessario individuare gli attori che ne rimangono coinvolti.

E' possibile schematizzare tre categorie di attori e le necessità che hanno nel momento in cui decidono di utilizzare RSS:

| Attore | Bisogno |
|---|--|
| fornitori di contenuti (publishers and content providers) | divulgare ad un pubblico più ampio possibile i propri contenuti |
| produttori di siti WEB | integrare sui propri siti WEB più informazioni possibili circa un dato argomento |
| utenti finali | accedere facilmente alle informazioni soprattutto nel momento in cui ci sono degli aggiornamenti |

Analizziamo ora i vantaggi per ciascuna categoria di attori:

- fornitori di contenuti.

In genere un utente che naviga su Internet vorrebbe ricevere informazioni dai propri siti preferiti, ma spesso non ha tempo o non si ricorda di navigarli. Se però l'utente ha installato sul proprio computer uno specifico programma detto "aggregatore" (in gergo anche "news aggregator" o FEED Reader) collegato ai FEED RSS dei siti di interesse, verrà automaticamente avvisato degli aggiornamenti avvenuti sul sito stesso. Questo aumenta la probabilità per quel sito WEB di essere visitato.

Un meccanismo simile potrebbe essere attivato mediante delle newsletter in cui agli utenti che si registrano su determinate aree del sito viene inviata una email ogni volta che avviene un aggiornamento di tali aree. Il meccanismo delle newsletter presenta però degli inconvenienti. Per il fornitore l'inconveniente più costoso si ha nella gestione e manutenzione di un servizio di mailing su grossi bacini di utenza; inoltre è necessario che gli utenti si registrino in un database e quindi ciò rappresenta spesso un deterrente per quegli utenti che per ragioni di riservatezza non hanno intenzione di lasciare nei vari siti WEB il proprio indirizzo email. Inoltre, per l'utente che riceve la newsletter risulta scomodo dover ogni volta ricercare all'interno della propria posta tale newsletter senza confonderla con posta spazzatura (spamming);

- produttori di siti WEB.

Il più grosso vantaggio per questa categoria è quello di usufruire del data sharing. In pratica attraverso l'uso di RSS è possibile integrare all'interno dei propri contenuti informazioni provenienti da altri siti senza dover duplicare i dati sul proprio sito;

- utenti finali.

Il vantaggio riscontrato dagli utenti finali è la possibilità di ricevere in tempo reale l'avviso di aggiornamento di un determinato sito senza dovervi preventivamente accedervi. Tale vantaggio potrebbe essere simile a quello offerto dai browser nella voce di menù "Preferiti". Però a differenza dei "news aggregator" l'utility implementata attraverso la voce di menu "Preferiti" fornisce solo la possibilità di memorizzare degli indirizzi di siti WEB in cui ci sono argomenti d'interesse per l'utente senza che ciò comporti una segnalazione nel momento in cui avvengono degli aggiornamenti in tali siti.

2.2 I probabili scenari di RSS

Sono molte le possibili implicazioni di questa tecnologia. La diffusione di RSS in Istat porterebbe notevoli vantaggi, in ambito comunicativo, sia verso l'esterno che verso l'interno (intranet). In particolare tutte le informazioni organizzabili in liste potrebbero essere buone candidate come FEED RSS:

- news;
- comunicati stampa;
- annunci circa conferenze ed eventi;
- avvisi di reclutamento;
- software release;
- nuovi acquisti di biblioteca;
- ecc..

Inoltre RSS potrebbe essere sicuramente una valida alternativa ad applicazioni di posta elettronica in tutte quelle situazioni in cui ne viene compromesso il rendimento: l'email è pubblica, ed essendo raggiungibile facilmente, può anche essere riempita con posta "spazzatura" o con informazioni di dubbio interesse. In pratica non essendo un canale dedicato, nella nostra inbox riceviamo notizie di varia natura: le email di lavoro si mischiano con le newsletter, le proposte commerciali si alternano alla posta personale. La posta elettronica è un serbatoio e come tale tende a riempirsi e, quindi, va sottoposto a manutenzione e pulizia. L'RSS è invece un mezzo dedicato alla ricezione esclusiva della categoria di FEED prescelta: nessuno può raggiungerci se non è abilitato a farlo.

Per quanto riguarda l'organizzazione di RSS c'è da dire che anche se è buona norma realizzare un "master FEED" con una lista delle notizie più recenti su un sito WEB, è necessario suddividere i FEED candidati per aree di affinità.

Per esempio consideriamo un sito di una biblioteca che intende utilizzare RSS per informare gli utenti circa le nuove acquisizioni. In questo caso è opportuno realizzare più FEED RSS per tema: letteratura, storia, matematica, geografia, ecc.

2.3 L'implementazione del protocollo RSS

Il protocollo RSS viene implementato attraverso la pubblicazione di un file. Tale file utilizza un formato XML organizzato in maniera differente a secondo della versione di RSS che si va ad utilizzare. In ogni caso cercando di generalizzare, un FEED RSS consiste di un canale (channel) con i suoi attributi (es. Titolo, descrizione, data di creazione, ecc.) e da un certo numero di argomenti (item) con i propri attributi (titolo, descrizione, URL, ecc.). Uno schema di base del file XML rappresentante un FEED RSS potrebbe essere il seguente:

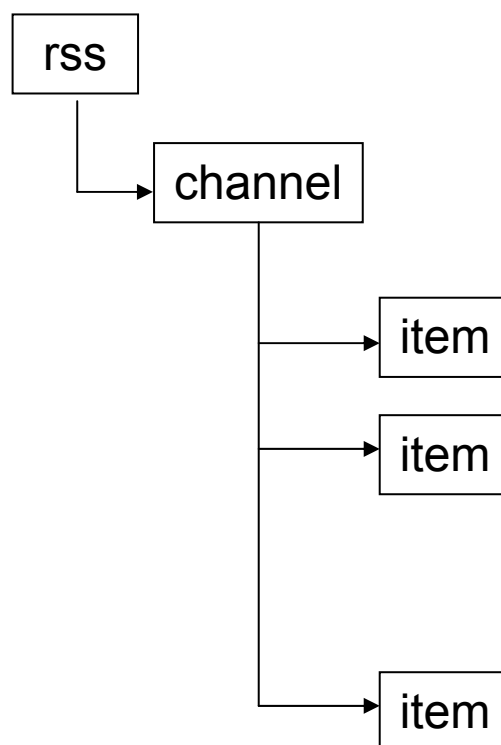


Figura 2.1

Per quanto riguarda le versioni del protocollo RSS è possibile dire che esistono due versioni maggiormente utilizzate: RSS 0.9x (e la sua evoluzione RSS 2.0) e RSS 1.0

La versione 1.0 infatti non rappresenta una evoluzione della versione 0.9x in quanto le due versioni principali rappresentano due tipologie di RSS differenti.

Di seguito si riporta uno schema riassuntivo delle varie versioni:

| Versione | Proprietario | Caratteristiche | Status |
|-------------------------|-----------------------|---|---|
| 0.90 | Netscape | | superata da 1.0 |
| 0.91 | UserLand | Molto semplice | Ufficialmente superata da 2.0, ma ancora utilizzata |
| 0.92, 0.93, 0.94 | UserLand | Permettono un inserimento di metadati più ricco rispetto a 0.91 | superata da 2.0 |
| 1.0 | RSS-DEV Working Group | Basato su RDF, estensibile attraverso moduli aggiuntivi, non controllato da una singola società | Modulo stabile |
| 2.0 | UserLand | estensibile attraverso moduli aggiuntivi, facilmente migrabile da 0.9x | Modulo stabile |

La versione RSS 2.0 che viene offerta dalla Berkman Center for Internet & Society at Harvard Law School² come "Attribution/Share Alike Creative Commons license", risulta compatibile con la versione 0.9x. Nel medio periodo è ragionevole pensare che molti dei siti web che utilizzano la versione 0.9x evolveranno alla 2.0.

All'interno dell'Istat ancora non sono state fatte delle scelte circa l'uso generalizzato dell'RSS: solo sulla homepage delle banche dati Conistat e Demo sono presenti dei file RSS, entrambi realizzati utilizzando la versione 2.0.

Segue un esempio di file in formato XML che utilizza il protocollo RSS 2.0. Non sono presenti tutti gli attributi in quanto come verrà descritto più avanti (vedi paragrafo 2.4) alcuni sono opzionali:

```
<?XML version="1.0" encoding="utf-8"?>
<RSS version="2.0">
  <channel>
    <title>calendario comunicati stampa</title>
    <copyright>(c) 2004, Istat</copyright>
    <item>
      <title>Fatturato altri servizi e commercio all'ingrosso</title>
      <pubDate>Wed, 30 Mar 2005 00:00:00 GMT</pubDate>
      <link>http://www.istat.it/comunicati/as_30_03_2005.pdf</link>
    </item>
    <item>
      <title>Fatturato altri servizi e commercio all'ingrosso</title>
      <pubDate>Fri, 01 Jul 2005 00:00:00 GMT</pubDate>
      <link>://www.istat.it/comunicati/as_01_07_2005.pdf</link>
    </item>
    <item>
      <title>Fatturato altri servizi e commercio all'ingrosso</title>
      <pubDate>Fri, 30 Sep 2005 00:00:00 GMT</pubDate>
      <link>://www.istat.it/comunicati/as_30_09_2005.pdf</link>
    </item>
  </channel>
</RSS>
```

² <http://cyber.law.harvard.edu/home/>


```

</item>
<item>
  <title>Fatturato altri servizi e commercio all'ingrosso</title>
  <pubDate>Fri, 23 Dec 2005 00:00:00 GMT</pubDate>
  <link>://www.istat.it/comunicati/as_23_12_2005.pdf</link>
</item>
</channel>
</RSS>

```

2.4 Struttura di un documento RSS

Il Protocollo RSS formalizza le regole da seguire per creare un documento RSS. In tale documento è previsto un nodo radice denominato “RSS” con un attributo “version” che specifica la versione di RSS che si va ad utilizzare. Segue il sottonodo <channel> con vari elementi e diversi sottonodi <item>, uno per ciascun occorrenza delle news che si vogliono mettere in evidenza.

I nodi presentano alcuni elementi obbligatori e altri opzionali. In seguito vengono descritti gli elementi ed i nodi principali previsti nella versione 2.0³.

2.4.1 Gli elementi del nodo channel

| Elementi obbligatori del nodo channel | | |
|---------------------------------------|---|---|
| Elemento | Descrizione | Esempio |
| title | Il nome del canale. Attraverso esso gli utenti si riferiscono al servizio | Calendario dei comunicati stampa |
| link | L'URL del sito WEB che corrisponde al canale | http://www.istat.it |
| description | Frase che descrive il canale | Il calendario dei comunicati stampa per l'anno 2005 |

| Alcuni elementi opzionali del nodo channel | | |
|--|---|-------------------------------|
| Elemento | Descrizione | Esempio |
| language | La lingua utilizzata | it-it |
| copyright | Specifiche di copyright | Copyright 2005, Istat |
| pubDate | La data di pubblicazione del contenuto nel canale. Tutte le date devono essere codificate secondo le specifiche RFC 822 | Sat, 07 Sep 1960 00:00:09 GMT |
| image | Specifica una immagine in formato GIF, JPEG o PNG che può essere visualizzata con il canale | |

³ Una descrizione completa della versione 2.0 è possibile consultarla all'indirizzo Internet: <http://blogs.law.harvard.edu/tech/rss>

2.4.2 Gli elementi dei nodi item

Un channel può contenere uno o più nodi item. Un item rappresenta una “notizia”, come in un giornale. Così la sua descrizione rappresenta un riassunto della notizia e il link punta alla notizia completa. In genere, nel caso un item contiene una descrizione completa e non riassuntiva della storia, si tende ad omettere il link. Tutti gli elementi dell’item sono opzionali, ma almeno uno tra gli elementi title o description deve essere presente.

| Alcuni elementi del nodo item | | |
|-------------------------------|---|--|
| Elemento | Descrizione | Esempio |
| title | Il titolo dell’argomento | Produzione industriale |
| link | Il link che permette di accedere all’argomento completo | http://con.istat.it |
| description | La descrizione dell’argomento | Produzione industriale (mese gennaio 1995) |
| pubDate | Data in formato RFC 822 di pubblicazione dell’item | Mon, 07 Feb 2005 00:00:09 GMT |

2.4.3 La generazione dei FEED RSS

Ci sono diversi modi per generare FEED RSS. Nel caso si decide di adottare un Content Management System (CMS) è importante verificare se contenga già procedure automatizzate per generare FEED RSS.

Nel caso ciò non sia previsto è possibile valutare le seguenti possibilità:

- self-scraping – rappresenta un modo abbastanza semplice per la realizzazione dei FEED. Esistono dei software, denominati scraping tool, che analizzano le varie pagine presenti sul sito WEB ed estraggono parti rilevanti da inserire nei FEED. Alcuni di questi tool utilizzano “espressioni regolari⁴” o espressioni XPath (XML Path Language) mentre altri richiedono una maggiore attenzione durante la realizzazione delle pagine HTML attraverso l’inserimento di opportuni tag;
- estrazione da repository – tale possibilità presuppone il fatto che alcuni metadati che si riferiscono ai vari contenuti vengano memorizzati in un database (repository). In questo caso è possibile realizzare delle procedure a valle dei processi di caricamento (subito dopo il caricamento oppure in momenti prestabiliti) che prelevando le informazioni necessarie dal repository creano i FEED RSS;
- scrittura manuale – tale possibilità presuppone un intervento di un operatore che utilizzando un editore di testo (nel caso di personale specializzato) o un software preparato per l’occasione, digiti le informazioni necessarie per i FEED.

⁴ Espressioni regolari: permettono attraverso la specificazione di parti di testi, parti di parole o di simboli, la ricerca di parole o testi all’interno di un documento.

Esistono dei programmi che sono in grado di verificare i FEED RSS. Evidentemente tali programmi vanno utilizzati nella messa a punto delle procedure automatiche o semi automatiche e, nel caso delle procedure manuali, tutte le volte che vengono realizzati i FEED.

2.4.4 I client RSS

Esistono principalmente due modi per utilizzare i FEED RSS:

- da parte degli utenti finali attraverso l'uso di alcuni programmi, detti aggregatori, che attivati all'accensione del computer, visualizzano attraverso delle icone la presenza di nuovi aggiornamenti presso i FEED sottoscritti;
- da parte di applicazioni, prevalentemente WEB, che leggendo i FEED li traducono in una forma presentabile e li integrano all'interno di un sito WEB.

2.4.4.1 Uso degli aggregatori

Forse il modo più facile per utilizzare FEED RSS è attraverso gli aggregatori⁵ molti dei quali sono scaricabili da Internet in versione free. Quasi tutti questi programmi presentano caratteristiche simili:

- definizione dei FEED d'interesse attraverso l'inserimento dell'indirizzo URL;
- definizione della modalità di refresh per ciascun FEED (ogni quanto tempo deve avvenire);
- possibilità di organizzare i FEED in sottocartelle;
- avviso attraverso finestre di popup e, nel caso di presenza di scheda audio, attraverso l'emissione di un particolare suono nel momento in cui viene verificato un aggiornamento.

In genere tutti i programmi aggregatori presentano una schermata suddivisa in tre frame: il frame di sinistra è il "FEED selector". Visualizza i titoli di tutti i FEED sottoscritti. In questo frame è possibile una organizzazione in cartelle. Il frame superiore destro è il "FEED display". Nel momento in cui si seleziona un particolare FEED, in questo frame vengono visualizzati i vari item di quel particolare FEED. Il frame inferiore destro è il "content display" in cui viene visualizzato il contenuto di un item.

⁵ Alcuni indirizzi Internet in cui è possibile trovare delle versioni free degli aggregatori sono i seguenti: <http://www.feedreader.com/> e <http://www.sharpreader.net/> (per Windows); <http://ranchero.com/downloads/NetNewsWireLite1.0.8.dmg.gz> (per Mac OS X), <http://www.disobey.com/amphetadesk/> (per Windows, Linux, Mac OS X).

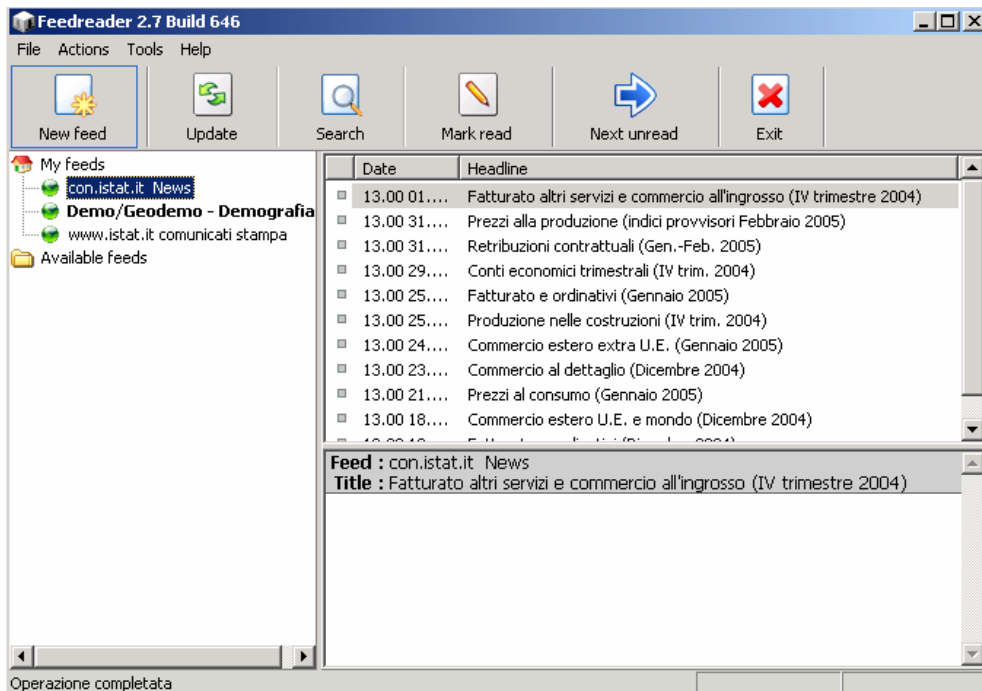


Figura 2.2

2.4.4.2 Applicazioni WEB che utilizzano FEED RSS

I FEED RSS possono essere utilizzati in tecniche di data sharing. Lo sharing di dati tra sistemi diversi richiede un formato di dati “neutrale” che possa essere facilmente trasmesso attraverso i protocolli Internet. XML si è dimostrato adatto allo scopo. Poiché XML è essenzialmente un file testo con un sistema di codifica ben nota, e poiché esistono parser XML facilmente utilizzabili da vari programmi e da linguaggi di programmazione, allora XML può essere utilizzato su qualunque sistema. RSS non è altro che una particolare grammatica XML.

In questo caso le applicazioni WEB devono attuare le dovute trasformazioni per rendere leggibili i file RSS. Ciò può avvenire in diversi modi. Il modo più semplice è quello di realizzare una trasformazione in formato HTML mediante un file XSLT (eXtensible Stylesheet Language Transformation):

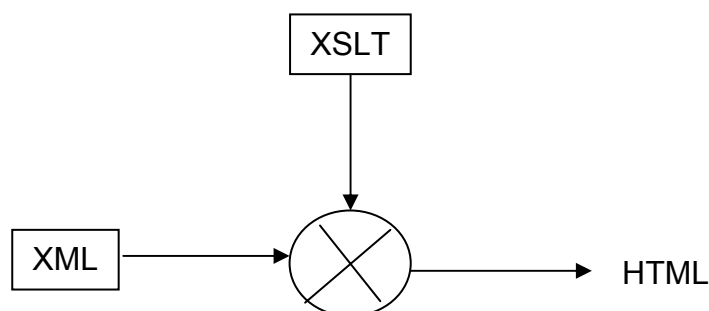


Figura 2.3

Come esempio consideriamo il seguente file comunicati.XML che rappresenta un FEED RSS dei comunicati stampa:

```
<RSS version="2.0">
  <channel>
    <title>www.istat.it comunicati stampa</title>
    <language>it-it</language>
    <link>http://www.istat.it</link>
    <description>comunicati stampa</description>
    <copyright>(c) 2005, Istat</copyright>
    <pubdate>Fri, 01 Apr 2005 15:39:21 GMT</pubdate>
    <item>
      <title>Stima preliminare del PIL - IV trimestre 2004</title>
      <description>Sulla base delle informazioni finora disponibili,
        nel quarto trimestre del 2004 il PIL, valutato ai
        prezzi del 1995 destagionalizzato e corretto per
        il diverso numero di giorni lavorativi, è
        diminuito dello 0,3 per cento rispetto al
        trimestre precedente ed è aumentato dell'1,0 per
        cento rispetto al quarto trimestre del 2003.
      </description>
      <link>http://www.istat.it/Comunicati/0404fsh.pdf</link>
    </item>
    <item>
      <title>Indice dei prezzi alla produzione dei prodotti
        industriale - Febbraio 2005</title>
      <description>L'Istituto nazionale di statistica comunica che,
        sulla base degli elementi finora disponibili, nel
        mese di febbraio 2005 l'indice generale dei prezzi
        alla produzione dei prodotti industriali con base
        2000=100 è risultato pari a 109,0, con un aumento
        dello 0,3 per cento rispetto al mese precedente e
        un aumento del 4,7 per cento rispetto al mese di
        febbraio 2004.
      </description>
      <link>http://www.istat.it/Comunicati/comp0205.pdf</link>
    </item>
  </channel>
</RSS>
```

Il seguente file comunicati.XSLT permetterà di fare la trasformazione in HTML:

```
<?XML version="1.0" encoding="UTF-8"?>
<xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform" version="1.0">
  <xsl:output method="html"/>
  <xsl:template match="/">
    <html>
      <head>
        <title><xsl:value-of select="//RSS/channel/title"/></title>
        <LINK href="styles.css" type="text/css" rel="stylesheet" />
      </head>
      <body>
        <table width="100%" border="1" bgcolor="gainsboro"
          bordercolor="white" cellpadding="0" cellspacing="0">
          <tr><td class="txt"><xsl:value-of
            select="//RSS/channel/title"/>
          </td></tr>
        </table>
        <table cellspacing="0" cellpadding="0" border="0"
          bordercolor="red" width="100%">
          <xsl:for-each select="//RSS/channel/item">
```

```

        <xsl:variable name="link" select="@link"/>
        <tr><td><A href="{\$link}"><xsl:value-of
            select="@title"/></A>
        </td></tr>
        <tr><td><xsl:value-of select="@description"/></td></tr>
    </xsl:for-each>
</table>
</body>
</html>
</xsl:template>
</xsl:stylesheet>

```

La trasformazione è stata fatto lato server utilizzando un piccolo programma scritto in Visual Basic .NET con application server ASP 1.1 e WEB server IIS 6.0 (vedi Appendice A)

Il risultato che si ottiene è il seguente:

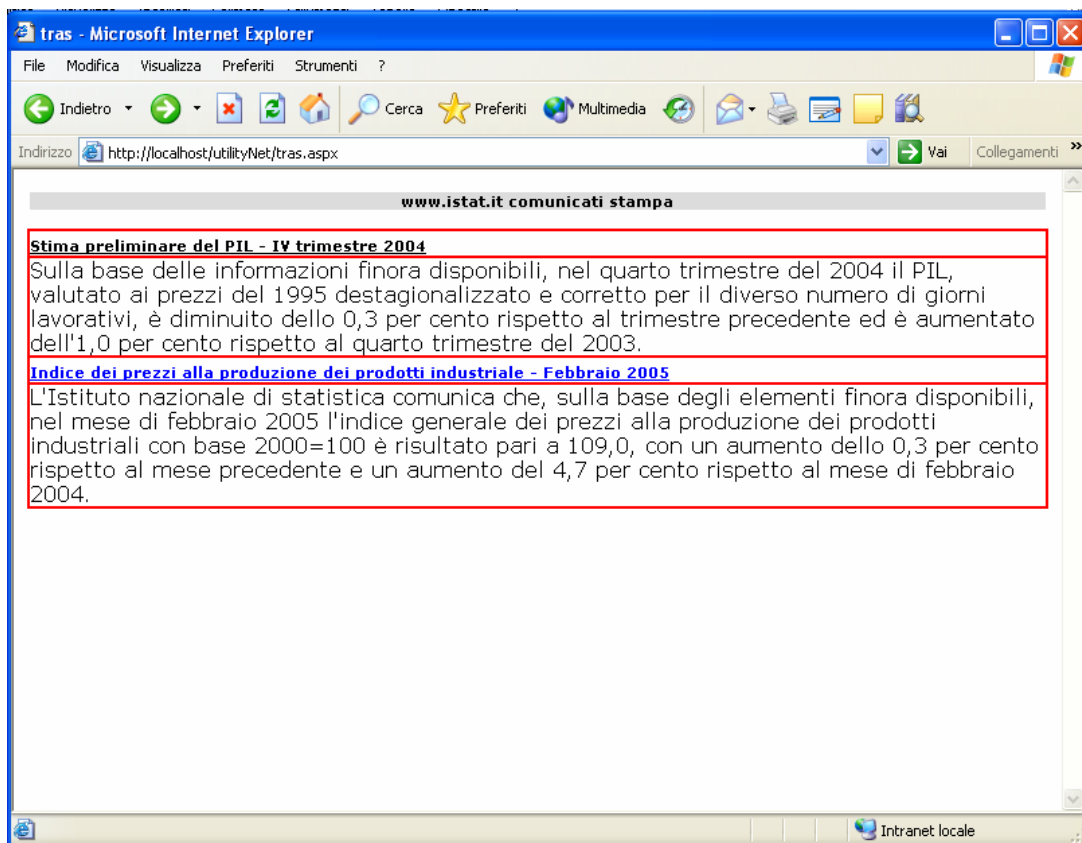


Figura 2.4

3. Due applicazioni sviluppate

All'interno della Direzione Centrale per la diffusione della cultura e dell'informazione statistica (DCDS) sono state sviluppate due applicazioni che danno luogo a FEED RSS. La prima, già in produzione, permette di realizzare un FEED RSS, sia in italiano che in inglese, a seguito di un aggiornamento della banca dati Conistat. La seconda, in fase prototipale, rappresenta un "editore" per inserire item corrispondenti ai comunicati stampa. Come è stato precedentemente detto (vedi paragrafo 2.4.3), esistono differenti modi per la realizzazione di FEED. I due software realizzati rientrano rispettivamente nel gruppo "estrazione da repository" e nel gruppo "scrittura manuale".

3.1 I FEED RSS per la banca dati Conistat

La banca dati Conistat contiene in prevalenza dati congiunturali che vengono aggiornati mensilmente dai Servizi di produzione secondo un calendario definito a inizio anno. Gli aggiornamenti, soprattutto in fase di periodiche revisioni dei dati, possono avvenire in date non considerate nel calendario. La homepage della banca dati contiene già un "banner" a scorrimento verticale che evidenzia gli aggiornamenti che vengono effettuati nella banca dati. Conseguentemente un utente che vuole vedere quali aggiornamenti sono stati effettuati, deve accedere alla homepage. E' stata quindi realizzata una procedura automatizzata per la produzione di due file, uno in italiano l'altro in inglese, che rappresentano due FEED RSS. La banca dati viene aggiornata attraverso una procedura client-server. A valle dell'aggiornamento, in automatico, una procedura provvede sia all'aggiornamento del banner delle news che dei file RSS.

In questo caso, quindi, il software che realizza i FEED rientra nel gruppo identificato come "estrazione dal repository".

Per la creazione dei FEED RSS è stato realizzato un software, scritto in MS Visual Basic 6.0, come modulo aggiuntivo al programma che aggiorna la banca dati. Il modulo software utilizza la libreria Microsoft MSXML v. 3.0 appositamente pensata per trattare dati in formato XML attraverso il modello DOM (Document Object Model)⁶.

In appendice B vengono riportati i dettagli tecnici, gli oggetti utilizzati e le loro caratteristiche.

⁶ Per una spiegazione circa il modello DOM vedere il link: <http://www.w3.org/DOM/>

3.2 Prototipo di uno strumento per scrivere FEED RSS

In questo caso il sistema realizzato per la generazione dei FEED rientra nel gruppo identificato come “scrittura manuale”, ovvero è stato realizzato un software ad hoc in cui l’utente inserisce direttamente i contenuti del FEED. Scopo del software realizzato è stato quello di sostituire con una procedura manuale l’impossibilità di prelevare in automatico, da un repository, le informazioni necessarie al FEED.

In pratica tale sistema è stato pensato per creare un FEED per segnalare l’avvenuta pubblicazione dei comunicati stampa.

La tecnologia utilizzata è Microsoft ASP.NET 1.1. Il vantaggio nell’utilizzo di tale tecnologia è nella facilità di creazione e gestione di file XML a partire da strutture di tipo relazionale.

Segue una descrizione dell’uso del software realizzato, mentre i dettagli tecnici sono illustrati nell’appendice C.

Per accedere al sistema è necessario autenticarsi fornendo una userid e una password. Tale operazione oltre a proteggere l’applicazione, consente di selezionare gli operatori in base alle loro competenze. In pratica gli operatori potranno accedere alle informazioni e alle funzionalità a cui sono abilitati.

La schermata che si presenta all’utente contiene i campi Titolo, Descrizione e Link relativi al comunicato stampa



The screenshot shows a web form titled "RSS - Comunicati" with the Istat logo. It contains two columns of input fields. The left column has labels "Titolo:", "Descrizione:", and "Link:" in red. The right column has labels "Title:", "Description:", and "Link:" in red. Below the fields is a "Verifica" button.

| Titolo: | Title: |
|---|---|
| Produzione industriale | Industrial production |
| Descrizione: | Description: |
| Produzione industriale (Febbraio 2005) | Industrial production (February 2005) |
| Link: | Link: |
| http://www.istat.it/Comunicati/In-calenda/Allegati/Impre | http://www.istat.it/Comunicati/In-calenda/All |

Verifica

Figura 3.1

Premendo il pulsante “Verifica” vengono effettuati due tipi di controlli:

- se i campi obbligatori sono stati effettivamente riempiti;
- se l’indirizzo specificato nel campo “Link” effettivamente esiste.

In assenza di errori apparirà la seguente schermata:

| | | | |
|---|---|---------------------|---|
| Titolo: | <input type="text" value="Produzione industriale"/> | Title: | <input type="text" value="Industrial production"/> |
| Descrizione: | <input type="text" value="Produzione industriale (Febbraio 2005)"/> | Description: | <input type="text" value="Industrial production (February 2005)"/> |
| Link: | <input type="text" value="http://www.istat.it/Comunicati/In-calenda/Allegati/Impre"/> | Link: | <input type="text" value="http://www.istat.it/Comunicati/In-calenda/Allegati/Imp"/> |
| <input type="button" value="Verifica"/> <input type="button" value="Salva"/> <input type="button" value="Nuovo"/> | | | |

Italiano

| | titolo | descrizione | link |
|---|---------------------------|--|---|
|  | Retribuzioni contrattuali | Retribuzioni contrattuali (Gen.-Feb. 2005) | http://www.istat.it/Comunicati/In-calenda/Allegati/Lavoro/Retribuzio/Rc050102.pdf |
|  | Produzione industriale | Produzione industriale (Febbraio 2005) | http://www.istat.it/Comunicati/In-calenda/Allegati/Imprese/indice-del/Pm0501.pdf |

Inglese

| | title | description | link |
|---|-----------------------|---|---|
|  | Contractual wages | Contractual wages (January-February 2005) | http://www.istat.it/Comunicati/In-calenda/Allegati/Lavoro/Retribuzio/Rc050102.pdf |
|  | Industrial production | Industrial production (February 2005) | http://www.istat.it/Comunicati/In-calenda/Allegati/Imprese/indice-del/Pm0501.pdf |

Figura 3.2

In tale schermata vengono mostrati tutti i comunicati stampa (sia in inglese che in italiano) presenti nel file RSS prima che venga modificando, più gli ultimi appena inseriti. Premendo il pulsante “Salva” il file RSS verrà aggiornato non riportando più le righe in cui l’icona RSS risulta barrata.

Il pulsante Nuovo consentirà l’inserimento di un nuovo comunicato stampa riportando l’utente alla schermata di inserimento (Figura 3.1).

APPENDICE A

Trasforma.aspx

```
<%@ Page Language="vb" AutoEventWireup="false" Codebehind="trasforma.aspx.vb"
      Inherits="utilityNet.tras"%>
<HTML>
  <HEAD>
    <title>trasformazione XML to HTML through XSLT</title>
    <meta name="GENERATOR" content="Microsoft Visual Studio .NET 7.1">
    <meta name="CODE_LANGUAGE" content="Visual Basic .NET 7.1">
    <meta name="vs_defaultClientScript" content="JavaScript">
  </HEAD>
  <body>
    <form id="Form1" method="post" runat="server">
      <asp:XML id="XML1" runat="server"></asp:XML>
    </form>
  </body>
</HTML>
```

Trasforma.aspx.vb

```
Imports System.XML
Imports System.XML.Xsl
Imports System.XML.XPath
Imports System.XML.Xsl.XslTransform
Imports System.io
Imports System.Text
Public Class tras
  Inherits System.WEB.UI.Page

  Private Sub Page_Load(ByVal sender As System.Object, ByVal e As
                        System.EventArgs) Handles MyBase.Load
    XML1.TransformSource = "xslt/comunicati.xslt"
    XML1.DocumentSource = "files/RSSIT.XML"
  End Sub
End Class
```

APPENDICE B

Di seguito vengono riportati i dettagli tecnici su come è stato realizzato il software per la generazione dei FEED di Conistat

La libreria di oggetti utilizzata è stata MSXML v. 3.0. Segue uno schema che relaziona gli oggetti del modello DOM con quelli della libreria:

| Oggetti DOM | Oggetti MSXML |
|--------------------------|------------------------------|
| DOMDocument | DOMDocument |
| DOMElement | IXMLDOMElement |
| DOMAttr | IXMLDOMAttribute |
| DOMProcessingInstruction | IXMLDOMProcessingInstruction |

L'oggetto DOMDocument rappresenta il documento XML con tutta la sua struttura formata da elementi, attributi, commenti etc..

L'oggetto DOMElement rappresenta un qualsiasi nodo del documento XML. Le specifiche XML stabiliscono che tra tutti i nodi che verranno definiti ne dovrà esistere uno principale che rappresenti il nodo radice. Nel nostro caso l'elemento radice è definito tramite il tag "RSS".

L'oggetto DOMAttr rappresenta un qualsiasi attributo che si può associare a ciascun oggetto DOMElement.

L'oggetto DOMProcessingInstruction rappresenta un nodo "processing instruction" contenente istruzioni di elaborazione. In questo nodo viene specificato il target "XML", la versione dell'XML e l'encoding⁷.

Per prima cosa viene creato l'oggetto Documento, quindi, procedendo come in una struttura ad albero, vengono creati gli altri oggetti e inseriti nel documento con i metodi AppendChild (che inserisce un nodo figlio all'interno di un nodo già esistente) e SetAttribute (che assegna un attributo ad un nodo già esistente).

Segue un estratto del codice utilizzato per la creazione del FEED RSS:

```
Dim cn1 As IXMLDOMElement, tlt As IXMLDOMElement
Dim lnk As IXMLDOMElement, descri As IXMLDOMElement
Dim lang As IXMLDOMElement, cprgt As IXMLDOMElement

Set cn1 = mReq.createNode(NODE_ELEMENT, "channel", "")
cn1.Text = ""
Call mroot.appendChild(cn1)
```

⁷ Attraverso l'encoding è possibile specificare il set di caratteri utilizzabili.

```

Set tlt = mReq.createNode(NODE_ELEMENT, "title", "")
tlt.Text = "con.istat.it News"
Call cnl.appendChild(tlt)

Set lnk = mReq.createNode(NODE_ELEMENT, "link", "")
lnk.Text = "http://con.istat.it"
Call cnl.appendChild(lnk)

Set descri = mReq.createNode(NODE_ELEMENT, "description", "")
If slng = "It" Then
    descri.Text = "Aggiornamenti di Conistat"
Else
    descri.Text = "Conistat updates"
End If
Call cnl.appendChild(descri)

Set lang = mReq.createNode(NODE_ELEMENT, "language", "")
If slng = "It" Then
    lang.Text = "it-it"
Else
    lang.Text = "en-en"
End If
Call cnl.appendChild(lang)

Set cprgt = mReq.createNode(NODE_ELEMENT, "copyright", "")
cprgt.Text = "(c) 2005, Istat"
Call cnl.appendChild(cprgt)
Do While Not f.AtEndOfStream
    strnews = f.ReadLine()
    data = Left(strnews, 10)
    descrizione = Trim(Right(strnews, Len(strnews) - 12))

    Set itm = mReq.createNode(NODE_ELEMENT, "item", "")
    Call cnl.appendChild(itm)

    Set itm_tlt = mReq.createNode(NODE_ELEMENT, "title", "")
    itm_tlt.Text = descrizione
    Call itm.appendChild(itm_tlt)

    Set itm_dt = mReq.createNode(NODE_ELEMENT, "pubDate", "")
    itm_dt.Text = CambiaData(data)
    Call itm.appendChild(itm_dt)
Loop

```

APPENDICE C

A differenza del caso evidenziato nell'appendice B, in cui il file XML viene realizzato come concatenazione di nodi in relazione padre-figlio, i FEED RSS per i comunicati stampa, vengono creati convertendoli direttamente da una struttura di tipo relazionale.

Il file XML che rappresenta il FEED RSS già esistente, viene caricato in un oggetto "dataset", tipico di ASP.NET. Ogni elemento del file XML nel dataset viene rappresentato da una tabella in cui le colonne rappresentano i nodi figli e gli attributi dell'elemento, collegate tra loro tramite un identificativo (chiave primaria e chiave esterna).

Il dataset, uno degli oggetti più utili di ASP.NET, permette l'accesso a qualunque fonte di dati in maniera disconnessa. Questo significa che è possibile collegarsi ad un database, prelevare i dati di cui si ha bisogno e chiudere la connessione verso il database⁸. Una volta che il dataset è stato riempito, è possibile riorganizzare, in memoria, tali dati secondo una nuova struttura relazionale. In pratica è possibile costruire nella RAM dell'elaboratore una struttura del tutto simile ad una database relazionale senza che essa sia presente fisicamente su disco.

Il dataset presenta alcuni metodi che permettono in maniera semplice di collegarsi a una fonte dati di tipo XML.

Nel caso in esame il flusso delle operazioni che vengono eseguite dal programma sono le seguenti:

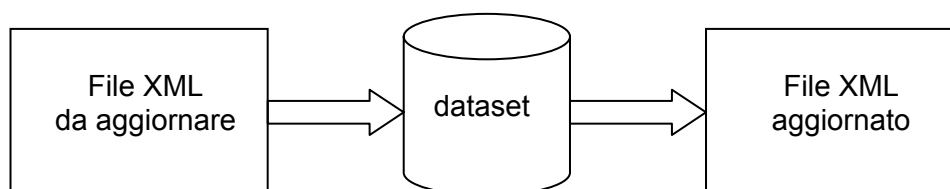


Figura A.1

Il caricamento del dataset da file XML comporta la creazione automatica di tre tabelle:

- una di nome "RSS" contenente le colonne "RSS_id" e "version";
- una di nome "channel", contenente le colonne "title", "language", "link", "description", "copyright", "pubdate", "channel_id", "RSS_id";
- una di nome "item" contenente le colonne "title", "description", "link", "channel_id".

⁸ La possibilità di lavorare con oggetti disconnessi aumenta molto la scalabilità di una applicazione soprattutto in ambiente WEB dove il numero di connessioni non è valutabile a priori.

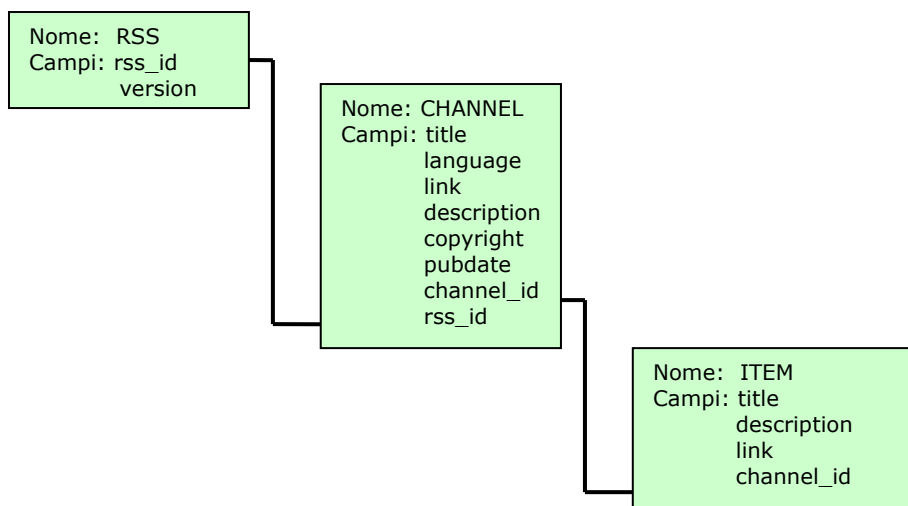


Figura A.2

Seguono le opportune elaborazioni per inserire nelle tabelle del dataset i record corrispondenti al nuovo comunicato stampa e controllare i record già presenti.

Nel caso in cui esistano dei record aventi data di pubblicazione più vecchia di quella corrispondente all'ultimo comunicato, verranno segnati, attraverso l'impostazione di un flag, come record da cancellare. Quindi viene creato un nuovo file XML aggiornato.

Per finire, attraverso un'operazione di collegamento dati (databind), viene utilizzato un oggetto "datagrid" che permette di inserire automaticamente in una tabella HTML i dati che si intendono presentare a video.

L'oggetto datagrid viene configurato opportunamente per offrire all'operatore una schermata esplicativa delle operazioni che ha compiuto.

APPENDICE D

Glossario:

| | |
|------------|---|
| XML | Extensible Mark-up Language. E' un linguaggio di markup basato su testo che fornisce informazioni di tipo strutturale e semantico relative ai dati . Questi dati sui dati, o metadati, offrono un contesto aggiuntivo all'applicazione che andrà ad utilizzare codesti dati e consente un nuovo livello di gestione e manipolazione dell'informazione basata su WEB. |
| XML schema | I file XML devono seguire due regole importanti: "buona formattazione" e "validità". Affinché tali regole vengano rispettate è necessario definire una grammatica. Un meccanismo per definire tale grammatica sono gli XML schema. |
| HTML | Hypertext Mark-up Language. Un formato di tipo testo che, attraverso un sistema di marcatori, descrive un file visualizzabile attraverso un browser WEB. |
| XSLT | eXstensible Stylesheet Language Transforms. E' uno standard che descrive come i documenti XML possono essere filtrati e convertiti in altri documenti XML. |
| Xpath | E' generalmente utilizzato insieme a XSLT per specificare quale parte del documento XML trasformare. XPath vede un documento XML come un insieme di nodi all'interno di un albero con relazioni di padre-figlio. Questi nodi possono essere di tipo differente: nodi elementi, nodi attributi, nodi di testo, ecc. Se XSLT può essere visto come un modo sofisticato di ricercare e trasformare una parte del documento XML, XPath è il modo per indicare quale nodo selezionare nella ricerca. |

Siti internet utili:

<http://w3c.org/>

World Wide Web Consortium

<http://www.rfc-editor.org/>

Request for Comments

<http://www.w3.org/DOM/>

Document Object Model (DOM)

<http://www.w3.org/XML/>

Sito XML del W3c

<http://www.w3.org/WAI/ut3/highlights/about-rss.html>

Sito RSS del W3c

<http://blogs.law.harvard.edu/tech/rss>

Specifiche RSS 2.0