Un viaggio Open Source nel semantic web

Federico Ruberti - Net7 Internet Open Solutions

Di cosa stiamo parlando...

"The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation."

Tim Berners-Lee, James Hendler, Ora Lassila, The Semantic Web, Scientific American, May 2001

Da dove veniamo...

★ WEB 1.0

La maggior parte degli utenti è read Only
45 Milioni di utenti globali (1996)
Focus sulle aziende
Home pages
Contenuti riservati
britannica online
HTML, Portali
Web Forms
Directories (taxonomy)
Netscape

★ WEB 2.0

Interazione continua tra lettura e scrittura
Più di un miliardo di utenti globali (2006)
Focus sulle community
Blogs
Contenuti condivisi
Wikipedia
XML, RSS
Web Applications
Tagging ("folksonomy")
Google

Web 3.0 - Il web dei dati

- ★ transformation of the Web from a network of **separately siloed applications and content repositories** to a **more seamless and interoperable whole**
- ★ ubiquitous connectivity, mobile Internet access and mobile devices;
- ★ **software-as-a-service** business models, Web services interoperability;
- ★ open technologies, **open data formats, open-source software platforms and open data** (e.g. Creative Commons, Open Data License);

Web 3.0 - Il web dei dati

- ★ open identity, open reputation, roaming portable identity ;
- ★ **Semantic Web technologies** such as RDF, OWL, SWRL, SPARQL, and statement-based datastores;
- ★ **distributed databases, the "World Wide Database"** ;
- ★ intelligent applications, natural language processing, machine learning, machine reasoning, autonomous agents.

Linked Data

Linked Data is about using the Web to connect related data that wasn't previously linked, or using the Web to lower the barriers to linking data currently linked using other methods. More specifically, Wikipedia defines Linked Data as "*a term used to describe a recommended best practice for exposing, sharing, and connecting pieces of data, information, and knowledge* on the Semantic Web using URIs and RDF."

Numerose esperienze:

- ★DBpedia - Semantic Enciclopedia
- ★Linked MDB - (Movie Data Base)
- ★Stumpedia - (Stumpedia.com is a social search engine that relies on human participation to index, organize, and review the world wide web.)

DBpedia

The DBpedia dataset describes 1,950,000 “things”, including at least:

- ★ 80,000 persons,
- ★ 70,000 places,
- ★ 35,000 music albums,
- ★ 12,000 films.
- ★ 657,000 links to images,
- ★ 1,600,000 links to relevant external web

DBpedia

The DBpedia dataset is interlinked on RDF level with various other Open Data datasets on the Web, like:

- ★ GeoNames
- ★ Musicbrainz
- ★ CIA World Fact Book
- ★ DBLP
- ★ Project Gutenberg
- ★ DBtune Jamendo
- ★ Eurostat
- ★ US Census data

Gestire conoscenza in evoluzione

- ★Scenario: contenuti e struttura stessa delle informazioni in rapida evoluzione

- ★Uno strumento di knowledge management necessita di upgrade frequenti in termini di:
  - Editing
  - Ricerca
  - Organizzazione (secondo diversi e nuovi "punti di vista")
  - Introduzione di nuovi concetti o relazioni tra essi

- ★Strumenti di knowledge management "Semantic Web Aware"
  - Netta separazione tra strato applicativo e struttura dell'informazione trattata

Vantaggi dell'approccio semantico

★…quindi:
- Riutilizzo e composizione di ontologie esistenti
- Modifiche strutturali delle ontologie a "run time"

  RDBMS "statici" vs. Ontologie "dinamiche"
- Funzionalità di editing e browsing che riflettono automaticamente questi cambiamenti strutturali

★In pratica:
- **Maggiore flessibilità rispetto all'evoluzione delle informazioni**
- **Minori costi di upgrade** (spesso non serve riscrittura di codice)
- **Riutilizzo degli stessi strumenti per diverse necessità**

Vantaggi dell'approccio semantico

★ Condividere e creare collaborativamente conoscenza semanticamente strutturata
  RDF facilita il "merging" di dati anche eterogenei
  URI come "punti di aggregazione" delle informazioni

★ Immediato riutilizzo/importazione di dati semantici
  – Da RDF pubblicati su Web
  – Da applicazioni Semantic Web esterne
  – Da flussi RSS
  – …

★ Pubblicazione su Web
  – Pubblicare RDF "crudo", immediatamente riutilizzabile da altre applicazioni
  – Rendere fruibili i contenuti su un portale semantico

Alcune applicazioni interessanti...

...per migliorare la produttività aziendale...

★Nepomuk (non è un nome Inuit...) - The social semantic desktop

- Networked Environment for Personal Ontology-based Management of Unified Knowledge -

(http://nepomuk.semanticdesktop.org/xwiki/bin/view/Main1/)

★Seek - Thunderbird extension

(http://code.google.com/p/simile-seek/)

...e dei tool di sviluppo completamente Open

- ★ Sesame - Framework for storage, inferencing and querying of RDF data
- ★ D2R server - Semantic Web server and SPARQL endpoint based on D2RQ that provides access to relational database content from within the Jena and Sesame RDF frameworks
- ★ Protege - Ontology editor

Per concludere...

- ★Web 1.0 - Open Source
- ★Web 2.0 - Open Access/Content
- ★Web 3.0 - Open Data (?)

Riferimenti slides

- Web 3.0 - DBpedia - Linked Data
  (relative voci su Wikipedia http://en.wikipedia.org/wiki/Main_Page)
- Gestione Conoscenza - Knowledge Management

  (Christian Morbidoni Sensible Logic - www.slideshare.com)

Federico Ruberti - ruberti@netseven.it

Net7 - Internet Open Solutions