

Conference paper

The Primary Care Electronic Library: RSS feeds using SNOMED-CT indexing for dynamic content delivery

Judas Robinson BA
Postgraduate Student

Simon de Lusignan MSc MD FRCGP
Senior Lecturer

Division of Community Health Sciences, St George's – University of London, UK

Patty Kostkova MSc PhD
Research Fellow, City eHealth Research Centre, Institute of Health Sciences, City University, London, UK

Bruce Madge FCILIP DHMSA PGCert HI
Library and Information Professional, British Medical Association Library, Tavistock Square, London, UK

A Marsh MSc PhD
Professor, University of Westminster, London, UK

C Biniaris PhD
Services Director, VMW Solutions Ltd, Romsey, Hampshire, UK

ABSTRACT

Background Rich Site Summary (RSS) feeds are a method for disseminating and syndicating the contents of a website using extensible mark-up language (XML). The Primary Care Electronic Library (PCEL) distributes recent additions to the site in the form of an RSS feed. When new resources are added to PCEL, they are manually assigned medical subject headings (MeSH terms), which are then automatically mapped to SNOMED-CT terms using the Unified Medical Language System (UMLS) Meta-thesaurus. The library is thus searchable using MeSH or SNOMED-CT. Our syndicate partner wished to have remote access to PCEL coronary heart disease (CHD) information resources based on SNOMED-CT search terms.

Objective To pilot the supply of relevant information resources in response to clinically coded requests, using RSS syndication for transmission between web servers.

Method Our syndicate partner provided a list of CHD SNOMED-CT terms to its end-users, a list which was coded according to UMLS specifications. When the end-user requested relevant information

resources, this request was relayed from our syndicate partner's web server to the PCEL web server. The relevant resources were retrieved from the PCEL MySQL database. This database is accessed using a server side scripting language (PHP), which enables the production of dynamic RSS feeds on the basis of Source Asserted Identifiers (CODEs) contained in UMLS.

Results Retrieving resources using SNOMED-CT terms using syndication can be used to build a functioning application. The process from request to display of syndicated resources took less than one second.

Conclusion The results of the pilot illustrate that it is possible to exchange data between servers using RSS syndication. This method could be utilised dynamically to supply digital library resources to a clinical system with SNOMED-CT data used as the standard of reference.

Keywords: computer communications networks, digital libraries, medical subject headings, primary care, SNOMED-CT, Unified Medical Language System

Introduction

Rich Site Summary (RSS) is a lightweight extensible mark-up language (XML) format for distributing and sharing content on the web; the process of sharing information is known as syndication. There are also other definitions of RSS: it has also been described as standing for Resource Description Framework (RDF) Site Summary and Really Simple Syndication.¹ Although most commonly used for news articles, RSS can be used to distribute any type of content. Hammond *et al* describe the role of RSS in science publishing;² PubMed (Medline) can deliver the results of database searches as RSS feeds.³ However, in this pilot we are interested in the syndication aspect of RSS, the process whereby a request from one web server returns content from another.

The source of the information content for this pilot is the Primary Care Electronic Library (PCEL). PCEL fits Witten and Bainbridge's definition of a digital library in that content can be added without altering the methods of access and retrieval.⁴ PCEL has 1500 indexed services, websites, databases and documents relating to primary care.⁵ Resources in PCEL are indexed using three methods: a custom classification unique to PCEL, medical subject headings (MeSH),⁶ and the Systematised Nomenclature of Medicine – Clinical Terms (SNOMED-CT).⁷ PCEL attracts between 500 and 1000 page requests per day, most during UK working hours. PCEL's technical specification is as follows: content is held in a MySQL database (version 4.1.14–standard),⁸ which is accessed using Hypertext Preprocessor (PHP) scripts (version 4.4.3)⁹ running on an Apache web server, which in turn sits on a Sun computer running the Solaris operating system. The database server also contains the 2005AB release of the Unified Medical Language System (UMLS).¹⁰ PCEL has provided RSS feeds for over a year¹¹ and also presents RSS content derived from diverse sources, principally news and medical news feeds.¹²

Our syndicate partner, Healthcare over IP (HoIP), is a community interest company set up on social enterprise principles to provide e-health solutions.¹³ HoIP wanted to provide its end-users with information about coronary heart disease (CHD). It wanted to use SNOMED-CT terms as key words to search for information and approached PCEL to deliver information relevant to its search terms. The HoIP end-user is intended to communicate via a portable device, which links to the HoIP web server. The HoIP web server uses different technology from PCEL – HoIP uses Active Server Pages (ASP) running on a Microsoft platform. This paper describes the feasibility of using SNOMED-CT terms to make requests from a remote

digital library and deliver content dynamically using RSS feeds.

Methods

We constructed an RSS feed using a standard method.¹⁴ The construction of the RSS feed is straightforward. Its components are: a title, a link to the URL of the relevant PCEL index card, a description and publication date. The PCEL index card is a web page displaying all the information and metadata concerning that resource. An example of the RSS feeds produced by PCEL is shown in Figure 1. The way information would be presented in the form of an index card is shown in Figure 2.

UMLS is critical to the functionality of PCEL. Firstly, we use UMLS to automatically map MeSH terms, which are manually applied to newly indexed content, to SNOMED-CT terms. Secondly, the UMLS generates important parts of the RSS feed produced by PCEL.

SNOMED-CT classification is achieved by mapping from MeSH to SNOMED-CT using the Metathesaurus of UMLS produced by the National Library of Medicine.¹⁵ UMLS classifies terms from numerous source vocabularies into concepts. These concepts are identified by Concept Unique Identifiers (CUIs). By identifying SNOMED-CT terms with the same CUIs as the indexed MeSH terms, mapping between the two vocabularies is possible. UMLS represents the cumulative efforts of many years of research by the National Library of Medicine.

We concatenated two strings present in UMLS (Path to Root [PTR] and Atom Unique Identifier [AUI]) to produce a new concept that we have named Hierarchical Unique Identifier (HUI). We did this to simplify the database structure. HUIs represent the polyhierarchical structure of SNOMED-CT. In the PCEL database, all terms are indexed by HUIs. Using the UMLS Metathesaurus we can map between the CODE and the HUI of a SNOMED-CT term. A list of information about resources can be produced on the basis of a CODE. The Metathesaurus is necessary to map between the CODEs that are supplied by the syndicated organisation and the HUIs contained in the PCEL database.

The RSS feed conveyed information about PCEL resources in response to the receipt of a Source Assorted Identifier (CODE) for a SNOMED-CT clinical term. The pilot concerned coronary heart disease and restricted itself to seven SNOMED-CT terms; these are shown in Figure 3. To give an example from the terms used in this pilot: the CODE for the

```

<item>
  <title>National Heart Forum (NHF)</title>
  <link>http://www.pcel.info/index.php?fuse=home.indexcard&id=645</link>
  <description>The National Heart Forum (NHF) is a leading alliance of over
  40 national organisations working to reduce the risk of coronary heart disease
  in the UK. The site discusses key facts and figures about the scale of the
  human, social and economic toll caused by coronary heart disease. There are
  links to related publications and information on Nutrition,Physical activity and
  Tobacco as ways of controlling CHD.</description>
  <pubDate>Thu, 06 Jan 2005 12:51:00 GMT</pubDate>
</item>
<item>
  <title>British Cardiac Society</title>
  <link>http://www.pcel.info/index.php?fuse=home.indexcard&id=708</link>
  <description>The British Cardiac Society was established in 1922, and is a
  charitable body. The BCS is the leading professional body for cardiology in
  the UK. The majority of our members are UK cardiologists and cardiac
  surgeons, as well as other doctors and healthcare professionals. The Society
  is involved in education, the setting of clinical standards and research into
  heart and circulatory diseases. This site is designed both for BCS members
  and for all those interested in cardiology.</description>
  <pubDate>Thu, 06 Jan 2005 10:43:00 GMT</pubDate>
</item>

```

Figure 1 Example CODE from an RSS feed

The screenshot shows the PCEL website interface. At the top, there is a search bar and navigation links: Submit, Site Map, Feedback, About, Help, Text only. The main header features the PCEL logo (a horse and rider) and contact information for St. George's University of London. Below the header is a navigation menu with links: Home, Directory, MeSH, SNOMED CT, EBM Search, Message Board, MyPCEL, News Feeds, Learning Log.

The main content area displays an index card for the National Heart Forum (NHF). The card includes a title, a description, and a list of related terms. The description states: "The National Heart Forum (NHF) is a leading alliance of over 40 national organisations working to reduce the risk of coronary heart disease in the UK. The site discusses key facts and figures about the scale of the human, social and economic toll caused by coronary heart disease. There are links to related publications and information on Nutrition,Physical activity and Tobacco as ways of controlling CHD."

Below the description, there is a table of metadata:

URL	http://www.heartforum.org.uk/
Cache	view cache
Organisation	NHF
Type	web site
Country	UK

Below the metadata, there are sections for Custom Directory, MeSH, and SNOMED CT, each with a list of related terms. The SNOMED CT section includes terms like "Disorder of the cardiovascular system" and "Primary prevention | Heart disease | Heart disease NOS | Disorder of cardiovascular system | [X]Cardiovascular disease, unspecified | Disease affecting entire cardiovascular system |".

At the bottom of the page, there is a footer with copyright information and links to Privacy policy, XHTML, CSS, pcel@sgul.ac.uk, Biomedical Informatics, and PCI.

Figure 2 Example index cards from PCEL

SNOMED-CT term 'Disorder of the cardiovascular system' is 49601007; one of the HUIs for the same term is A3684559.A3886745.A3456474.A3340519.A6938265. To map between these concepts, two tables in UMLS are used: MRCONSO (Concept Names and Sources) and MRHIER (Computable Hierarchies).

The MRCONSO table contains a line of data for every atom contained in the Metathesaurus, that is, it contains a line of data for every SNOMED-CT term. By querying MRCONSO the CODE value can be mapped to one or more Atom Unique Identifiers. Using the AUIs that are found, the MRHIER table

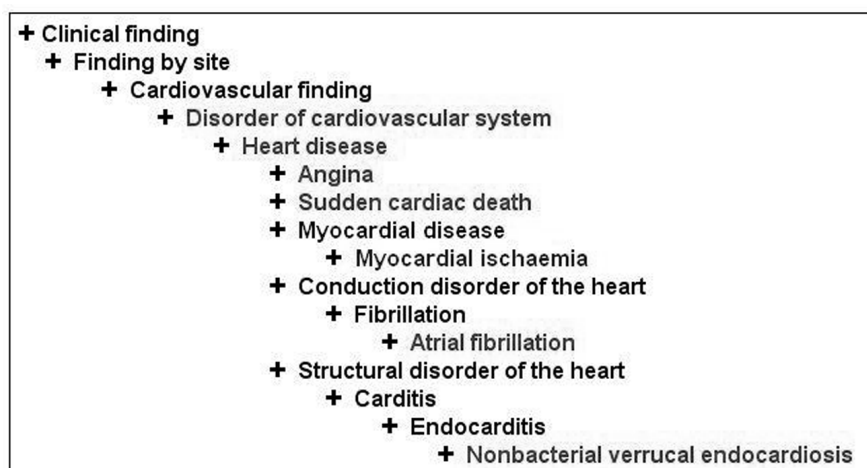


Figure 3 Hierarchical structure of SNOMED-CT

can be queried for the Path to Root values that are associated with them. By combining the PTR and the AUI we get the HUI, which is the value by which information about SNOMED-CT classifications is stored in PCEL.

The final step in the process is the utilisation of the RSS feed. Once our syndicate partner has requested the RSS feed corresponding to a given SNOMED-CT CODE, the XML data that composes the RSS feed is delivered via the HTTP protocol. Our syndicate partner uses Active Server Pages to parse the XML of the RSS feed and then presents the content online. It is important to note the platform independence of this process.

Results

The pilot demonstrates the ability of the utilised technologies to exchange data using RSS syndication in response to requests using SNOMED-CT terms. The system operates in real time, returning results in

less than half a second. The process is automated using look-ups in the UMLS Metathesaurus database. The pilot interface can be viewed online and is illustrated diagrammatically in Figure 4.

The HoIP system is designed to be used with WinCE operating systems (used for mobile devices). Unfortunately the system is not optimised for everyday internet browsers so does not display well when viewed through them. To get an idea of how the application works, the user needs to right-click on the links that scroll across the screen and select 'Open in new window'. Seven SNOMED-CT terms are displayed and by selecting these and refreshing, results from PCEL will be displayed. This process is entirely dynamic, so if the resources on PCEL are updated, the material returned will reflect this change.

The process delivered results according to the specification requested: that is, search results restricted to the specific single SNOMED-CT CODE. However, initial feedback from the pilot suggested that users might prefer to set their own searches either to be restricted to their search term or 'exploded' to capture information more widely held about the subject. For example, angina is a child concept of the term 'Heart

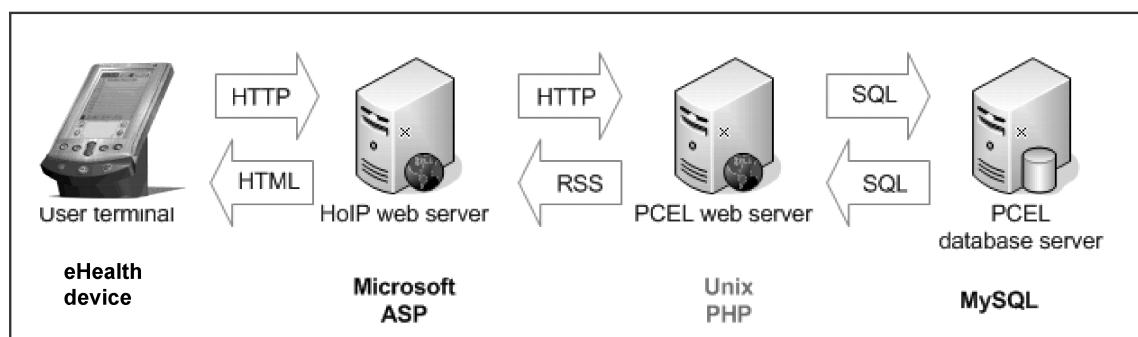


Figure 4 The architecture of the PCEL/HoIP pilot

Disease'. At the moment, when a user views the resources indexed under Heart Disease, they do not automatically see the resources for Angina. That is, the views of SNOMED-CT are not exploded. Perhaps it should be a matter of user preference whether or not the terms should be exploded: explosion might retrieve too many resources; alternatively focus might decrease sensitivity. The option to explode should be provided to users; the PCEL database stores data about SNOMED-CT terms in their hierarchical form, which means that this would be a simple technical feature to add.

Archived PCEL resources, not normally visible to the direct user of PCEL but remaining on its database, were returned to HoIP users. This is because resources in the PCEL database can be marked in order that they are not displayed. This functionality enables the withholding of websites whose URL is out of date, whose content is out of date, or is unsuitable for other reasons. This discrepancy between the pilot and the digital library is a minor problem and can be rectified when we move from the pilot to the live application.

Discussion

This pilot demonstrated the use of RSS syndication to distribute the contents of a digital library. Dynamic requests were produced on the basis of a SNOMED-CT term. The unique identifiers in UMLS linked a SNOMED-CT CODE to the polyhierarchical structure of SNOMED-CT represented as HUIs in our database, to the individual resources returned from the digital library.

Whilst RSS feeds are commonplace for the delivery of evolving content such as news, we are using the technology in a novel way. We have linked clinical terms to a Metathesaurus within PCEL and linked this to the appropriate resources.

The implications for practice are that it may be possible to link clinical data to appropriate information resources. Although the study was run on a small scale its components are scalable. PCEL only contains 1500 resources, whereas there are 700 000 clinical terms in SNOMED-CT. To produce a system that was capable of providing resources on the basis of granular SNOMED-CT clinical terms would require increasing the volume of indexed material on PCEL.

The solution to the problem of whether to restrict or explode requested hierarchies is to provide user choice. The breadth of the SNOMED-CT classification system might lead to problems when accessing a scaled version of the pilot. It is common that picking lists on the basis of search terms become unmanageable with larger terminologies when searching on basic terms such as 'Diabetes'. This might well be the case for a terminology containing 700 000 terms such as

SNOMED-CT. From our own experience such problems can be overcome with learned search strategies and familiarity with the classification system concerned.

A further challenge is the large number of synonyms that a given term might have. In part this difficulty is avoided by the polyhierarchical representation of SNOMED-CT in the PCEL database. The application will look for resources in all of the nodes of the hierarchy that are synonyms of each other, exclude repeated results and return a distinct subset of resources. Another potential clarifier for the application is its choice of Source Asserted Identifiers (CODEs) rather than Atom Unique Identifiers (AUIs) for communication between the two servers. There are only 300 000 CODEs compared with over 700 000 AUIs, and this reduction in the quantity of terms used will obviously reduce the number of synonyms encountered.

Despite these limitations, the study has potential applications. We have demonstrated the ability of a remote computer to request, via RSS feeds generated as the result of HTTP requests over the internet, resources relevant to a given SNOMED-CT term.

A critique of the mapping between the MeSH and SNOMED-CT is beyond the scope of this paper. However, whilst problems have been identified with this mapping, it is generally regarded as more than satisfactory.^{16,17}

Conclusions

Using RSS syndication, this pilot has delivered digital library content on request using SNOMED-CT Source Asserted Identifiers (CODEs). Were PCEL to be scaled in terms of the quantity of resources indexed, it would be possible to provide granular access to targeted internet resources on the basis of SNOMED-CT terms. There are advantages to this, given the national policy to implement SNOMED-CT across all health domains.

ACKNOWLEDGEMENTS

Thanks are due to all the staff at St George's, University of London. In particular, thanks are due to Dan Bird for setting up the UMLS Metathesaurus on a MySQL database server and to Paul Key for systems development. Also to the computer department for networking and creating order that made the whole project possible.

REFERENCES

- 1 Hammersley B. *Content Syndication with RSS* (1e). Sebastopol, CA: O'Reilly & Associates; 2003.
- 2 Hammond T, Hannay T and Lund B. The role of RSS in science publishing: syndication and annotation on the

- web. *D-Lib Magazine* December 2004; 10(12). www.dlib.org/dlib/december04/hammond/12hammond.html
- 3 Canese K. RSS feeds available from PubMed[®]. *NLM Technical Bulletin* 2005 May–Jun;(344):e2. www.nlm.nih.gov/pubs/techbull/mj05/mj05_rss.html
- 4 Witten I and Bainbridge D. *How to Build a Digital Library*. New York: Elsevier; 2002.
- 5 Robinson J, de Lusignan S and Kostkova P. The Primary Care Electronic Library (PCEL) five years on: open source evaluation of usage. *Informatics in Primary Care* 2005; 13(4):271–80.
- 6 National Library of Medicine. *Medical subject headings (MeSH)*. www.nlm.nih.gov/mesh/meshhome.html
- 7 SNOMED International. *SNOMED: SNOMED-CT*. www.snomed.org/snomedct/
- 8 MySQL AB. *MySQL AB: the world's most popular open source database*. www.mysql.com/
- 9 The PHP Group. *PHP: Hypertext Preprocessor*. www.php.net/
- 10 National Library of Medicine. *Unified Medical Language System (UMLS)*. www.nlm.nih.gov/research/umls/
- 11 Primary Care Electronic Library. *RSS feed of recently added sites*. www.pcel.info/pcel/rssmultiple.php
- 12 Primary Care Electronic Library. *News feeds*. www.pcel.info/index.php?fuse=home.news
- 13 Healthcare Over IP. www.hoip.org.uk/newsbar.html
- 14 XML.com. *What is RSS?* www.xml.com/pub/a/2002/12/18/dive-into-xml.html
- 15 Robinson J, de Lusignan S, Kostkova P and Madge B. Using UMLS to map from a library to a clinical classification: improving the functionality of a digital library. *Studies in Health Technology and Informatics* 2006;121: 86–95.
- 16 Caviedes JE and Cimino JJ. Towards the development of a conceptual distance metric for the UMLS. *Journal of Biomedical Informatics* 2004;37(2):77–85.
- 17 Cimino JJ, Min H and Perl Y. Consistency across the hierarchies of the UMLS Semantic Network and Metathesaurus. *Journal of Biomedical Informatics* 2003; 36(6):450–61.

This paper was presented at the PHCSG Annual Conference, Eynsham, Oxfordshire, September 2006.

CONFLICTS OF INTEREST

None.

ADDRESS FOR CORRESPONDENCE

Simon de Lusignan
Primary Care Informatics
Division of Community Health Sciences
St George's – University of London
London SW17 0RE
UK
Tel: +44 (0)208 725 5661
Fax: +44 (0)208 725 3584
Email: slusigna@sgul.ac.uk

Accepted November 2006