Smartlogic™

Case Study

Office of Public Sector Information (OPSI)

How OPSI bridged the terminology gap between experts and the public

The OPSI website (www.opsi.gov.uk)

The Office of Public Sector Information (OPSI) website provides access to Acts of Parliament and other UK legislation for the public, the information industry, government, and the wider public sector.

Working with OPSI since 1996, The Stationery Office (TSO) is responsible, under the authority of OPSI, for publishing enacted legislation and maintaining all legislation content on the OPSI website.

In an average month, the site receives nearly one and a half million unique visitors who generate almost thirteen million page impressions. It contains over 300,000 pages and over 55,000 documents. In 2008, Hitwise – Experian's web-traffic analysis firm – placed the OPSI site in the top ten UK central-government sites.

The website is substantial, it's heavily used, and it now has a sophisticated taxonomy-based classification and search-navigation solution to help users find relevant content.

The challenge: the terminology gap and the need for search-navigation features

Like many websites, the OPSI site is used by a wide range of people. Site users can be lawyers looking for recently published Acts of Parliament, information firms searching for licensing agreements, or citizens and small businesses who want to know about government policy on recycling. All of these different users will use different words and phrases when they're searching for information on the same subject.

For example, a lawyer specialising in EU-law might type in 'EU legislation' when searching for Acts of Parliament that enact EU directives, while a concerned citizen might tap in 'Brussels' (common shorthand for EU intervention in the UK) when looking for the same thing.

The terminology gap

We call this the terminology gap, and bridging it was OPSI's first challenge.

It means that an effective search solution has to look at different users' search-terms, interpret what subjects they're interested in, and then offer relevant results.

But this is only part of the terminology gap; the next problem is that website content is almost always written by expert authors. They tend to use specialist terms and formal language designed to leave no ambiguity, and as it's this content that's indexed by search engines, experts' language widens the terminology gap.

Search navigation; helping people find their way through the search process

The next challenge was to help people to make sense of large sets of search results. This would mean developing useful search-navigation features like concept maps, related topics, links to helpful resources and so on. Concept maps show users the topics that relate to their keyword searches, and helps users see how a subject area breaks down into topics.

A combination of search-navigation features and concept maps helps people both understand a subject, and learn more about it.

The solution: taxonomy-based classification brings improved 'findability'

The OPSI site uses Google Search Appliance (GSA), combined with Smartlogic's Semaphore. OPSI have bridged the terminology gap and improved 'findability' through a combination of:

- GSA's powerful keyword-search over huge volumes of content and familiar interface.
- Semaphore's taxonomy-based classification and search-navigation features.

Behind the scenes of the OPSI website, Semaphore automatically classifies any new content according to a taxonomy that meets e-government standards, and has been extended to accommodate the language of OPSI's different user types. As Semaphore classifies content, it adds metadata (invisible labels or 'tags') that identify the topics covered by an item of content.

TSO's workflows have been developed closely with OPSI to publish legislation content to the OPSI site on a daily basis in a way that compliments OPSI and Smartlogic's search solution. This includes the automated categorisation of every published legislation item using Smartlogic's Semaphore.

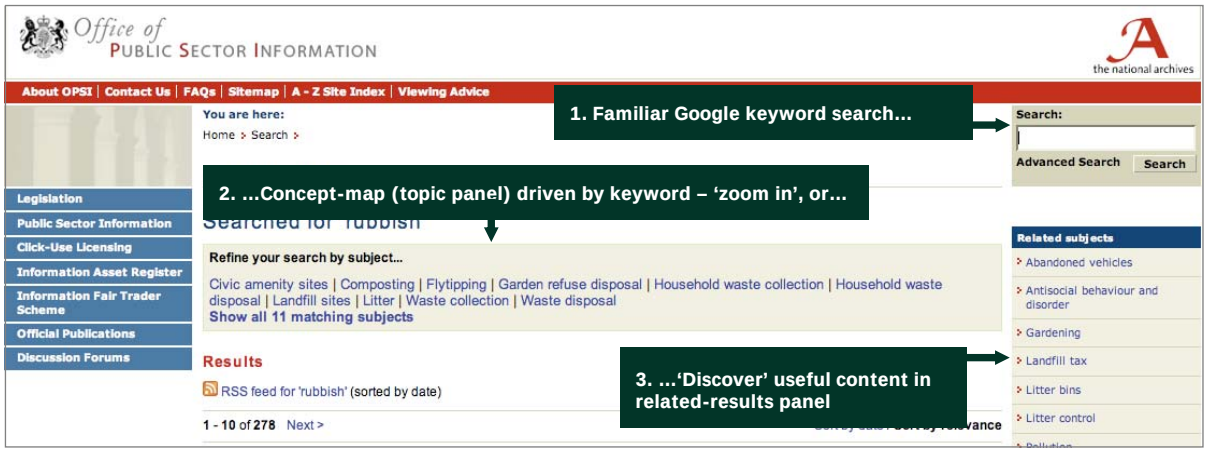
The taxonomy-driven tags are used in two important ways:

- Google Search Appliance uses the tags to increase the completeness of search results. The solution finds relevant documents even if they don't contain the user's search term (if you searched for 'manual digging implements' it would return results about spades).

- Semaphore uses the tags to drive the search-navigation features (see screenshots on following pages). For example, when you do a keyword search the search-results page shows a list of relevant and related topics, as well as a list of documents.

Search-navigation features help users 'zoom in' on, and 'discover' content

The search-navigation features on the OPSI website now display topics from a list of over 3,000.



These topics are a key part of the search-navigation experience and have been a hit with site users –in 60% of all queries made people used the navigation features to find what they were looking for.

In the screenshot above we've highlighted two search-navigation features that improve 'findability' on the OPSI site:

- *Concept maps that are driven by the keyword search*
  Do a keyword search, and a set of subjects that relate to that search appear in the beige topic-panel above the search results. This is a mini concept-map, and users can 'zoom in' on a topic by clicking on a word or phrase.

- *Search-navigation features that help users 'discover' more about a subject*
  Both the topic panel and the related-results feature help users learn more about a subject; they can 'discover' useful content by using the topic panel or the related-results list.

The decision: Google Search Appliance plus Semaphore

For John Sheridan, Head of e-Services at the UK Office of Public Sector Information, the combination of the Google Search Appliance and Smartlogic's Semaphore software delivered the best of both worlds:

"We knew that we'd need to combine keyword search with taxonomy-based classification, and search-navigation. We needed an easy-to-use search and navigation interface that would work for both expert and everyday users."

"Combining the Google Search Appliance (GSA) with Smartlogic's Semaphore product gave us the best of both worlds. We've got Google's powerful keyword search and familiar interface, and Smartlogic's taxonomy-based classification and search-navigation features."

Sheridan was also impressed by the work that Google and Smartlogic had done to knit the two products together, and still keep the ease of set-up that makes GSA so attractive:

"We took a long look at the GSA-plus-Semaphore offering, and everything that we saw gave us confidence. Smartlogic are a Google Enterprise Professional partner with a clever set of taxonomy tools, extensive experience of government taxonomies, and a neat bundle of search-navigation features. And Google's GSA has phenomenal search performance – particularly over large volumes of content – plus the set-up is so straightforward."

Smartlogic, Semaphore, and the Google Search Appliance

Smartlogic (the firm that develops and delivers Semaphore) was awarded "Most innovative Product Integration" at Google's Enterprise Professional Partner Awards in 2007.

Smartlogic's Semaphore software works closely with the Google Search Appliance, adding complete support for taxonomies, ontologies, automated classification and faceted search.

Semaphore helps organisations to:

- Enrich the Google search experience with the ‘findability’ benefits of search-navigation features.
- Rapidly build, import, and manage taxonomies and ontologies.
- Organise content according to a taxonomy by applying sophisticated rule-based analysis
  of the text – the most accurate and controllable approach to classification.
- Integrate multiple information sources and applications and make them available to users.
- Make content metadata from one system available to other applications and websites

Contact details

Smartlogic Semaphore Limited
Wellington House
East Road
Cambridge
CB1 1BH
United Kingdom

Phone: +44 (0)1223 451046
Fax: +44 (0)1223 451100

Web: www.smartlogic.com
Email: info@smartlogic.com