LexisNexis®

Patent Prior-Art Searching with Latent Semantic Analysis

Dr. Stuart McLean, LexisNexis®

Highlights

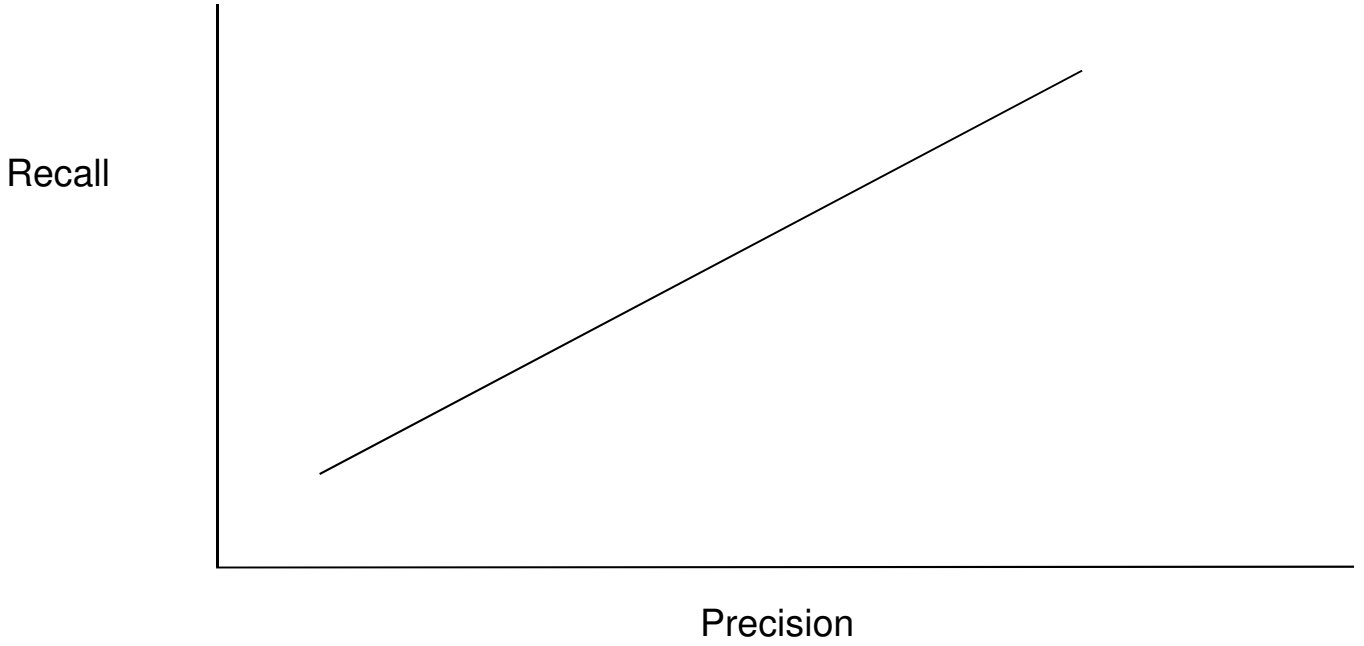
- ✓ Why we implemented Semantic Search
- ✓ What is Semantic Search
- ✓ The Process for Semantic Search in TotalPatent®
- ✓ UI Features

The Goal of Searching

"To leverage the subject matter expertise of the user so as to assist them in quickly and efficiently identifying all of the relevant art in a relevancy ranked manner and to eliminate the irrelevant art."

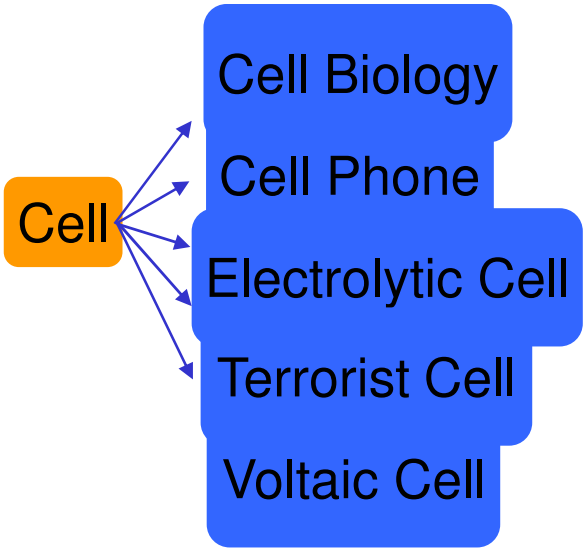
Our users require both:

- **Precision:** Retrieving a high level of accurate results relevant to your search query (a measure of exactness)
- **Recall:** Retrieving a high percentage of relevant documents (a measure of completeness)

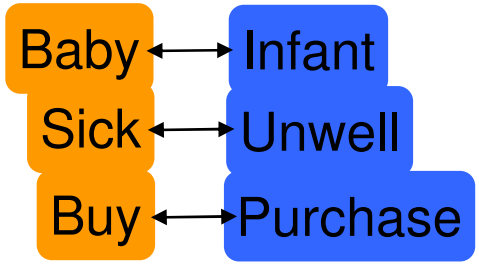


Why Context is Important

- **Polysemy** – Single words (or phrases) with multiple meanings, like:



- **Synonymy** - Multiple words with the same meaning, like:



Why is Patent Search unique?

Challenges Specific to Patent Searching:

- Disparate Nomenclature across art
- Applicants can act as own Lexicographers
- Industry language changes over time
- Technical abbreviations often conflict with common words (Au = Gold, Australia)
- Many technical abbreviations are common noise words (e.g. He = Helium, Be = Beryllium)
- Lengthy compound expressions are often misspelled
- 1000's of Certificates of Correction a year

Challenges with Search Systems

Search Intelligence is relatively static:

- State of the art still involves Lexicons, Synonym Lists and Thesauri.
- Semantic Enhancement identifies shared entities, not shared concepts
- Search suffers from chronic amnesia (user starts over on each search)

Search is unmanageable/disconnected:

- Too many search engines
- Too many indexes
- Too much disparate data

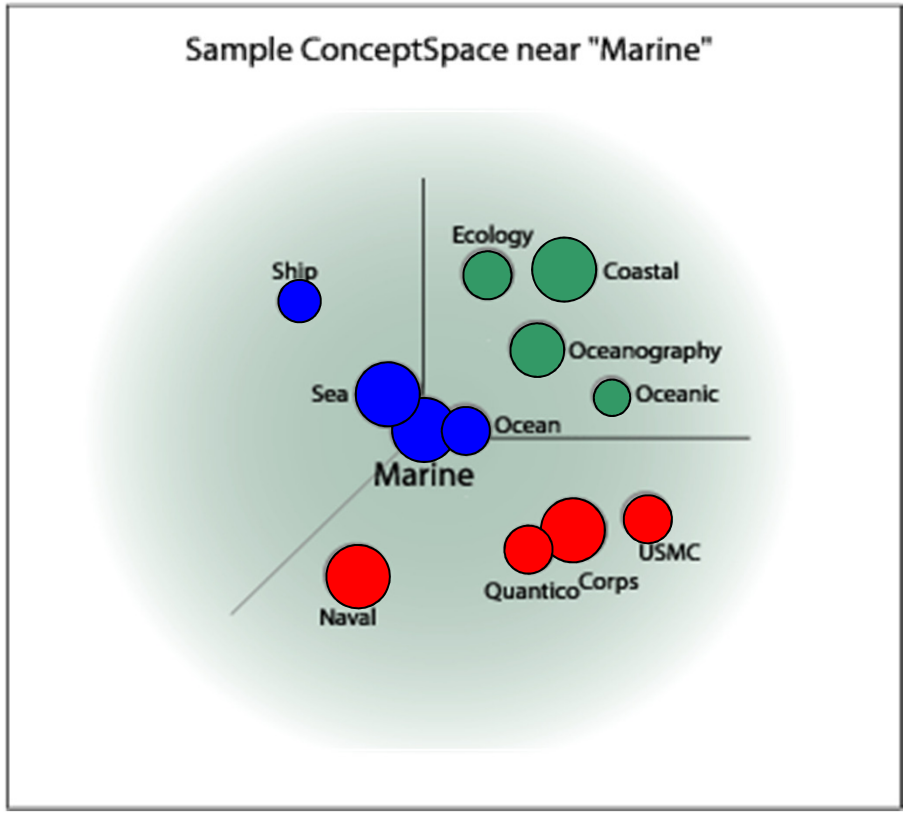
Search is Time Consuming:

Disparate, disconnected systems with no common language require too much time for one to be exhaustive and comprehensive – and efficient.

What is Semantic Searching

- **LSA extracts every contextual relationship among every word or phrase within a document set (8 million US Patents and over 3,000 STM titles). It then generates a vector space representation of all terms based on those relationships. Within that space, proximity is a strong indicator of conceptual similarity.**

- **<u>The result</u>: similarities can be identified based on concepts found within the collection itself.**



The Black Box - Surrendering Transparency, Control and Scale

TRANSPARENCY

- Semantic Search is effective and many times better than straight keyword searching, but thus far has locked users out of understanding **how** results are generated. We are forced to "trust" its algorithms, trust its function and trust its results. Virtually no system transparency.

CONTROL

- There is a surrendering of "control" with semantic search. Users cannot control how a query is constructed, much less the algorithms that create the search logic.

SCALE

- Semantic Searching requires a semantic index. Control of the search corpus via content indexing (and thus control over formatting, storage, etc.). This just does not scale. The web may never be semantically indexed.

The Result: Semantic searching is not a practical search alternative to traditional search methodology. It may even be a liability.

The Process

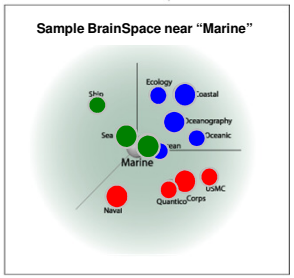
1) User builds Query

Semantic Query
* Words
* Sentence
* Paragraph
* Document,

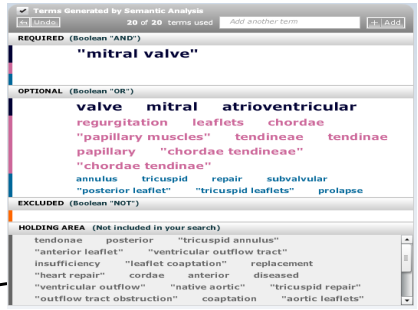
etc.

2) Brain is Selected (19)

3) Intelligence/Inference: Query terms extracted and inferred terms from the brain are added.



4) Transparency: QueryCloud is generated. User is shown everything generated by the machine learning.



Total time elapsed = 1 second

5) Automatic Query Generation: The System uses the words extracted and generated by the brain to formulate fully optimized queries for each index being searched.

Query Creation/Optimization

Simple Boolean

Nested Boolean

Ranked OR

Term Weights

TotalPatent®

Value & What it's Not

LexisNexis™

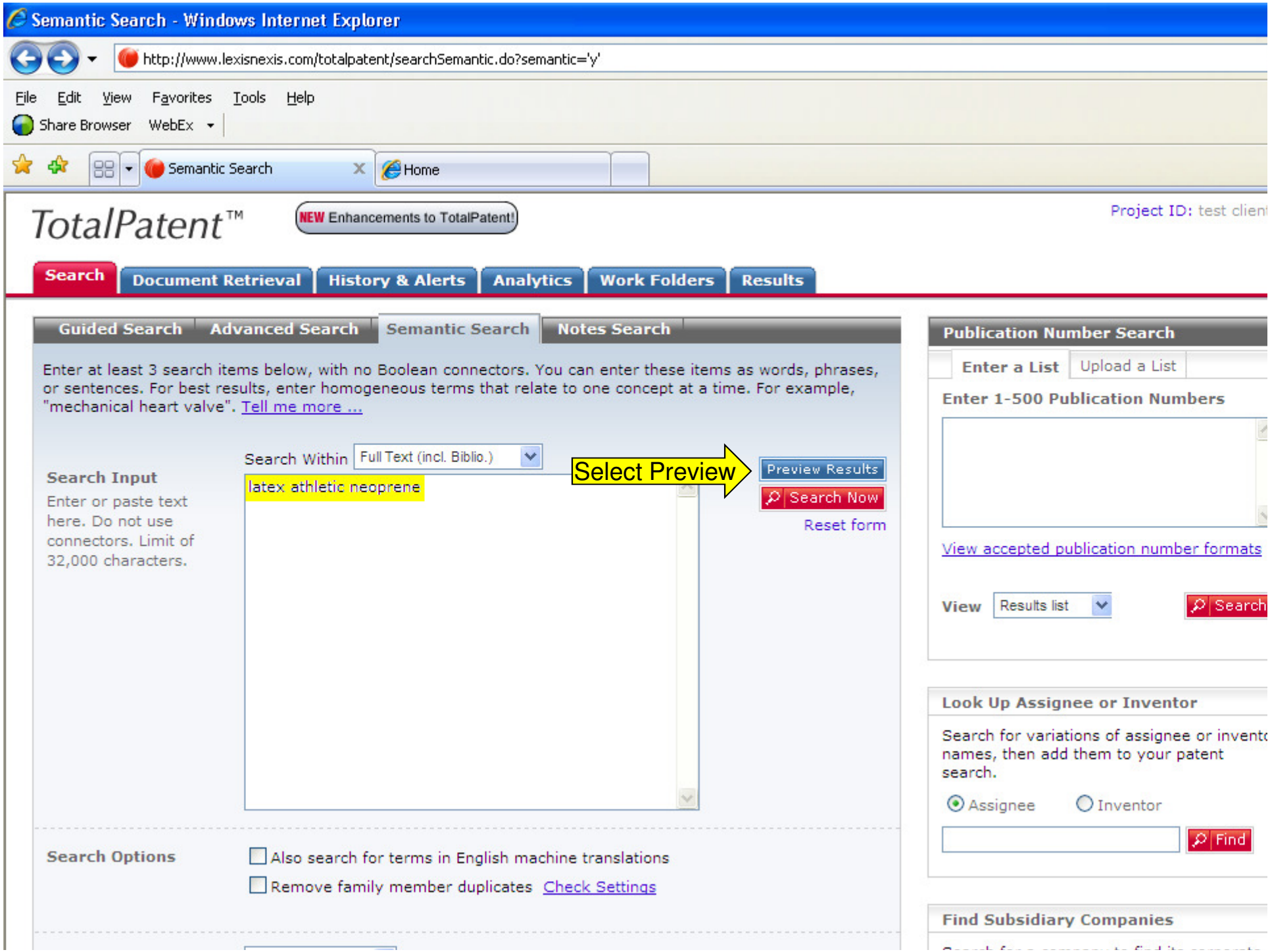
Why this is valuable?

- Full transparency into the results of the Semantic analysis
- Full enablement of Boolean tuning to allow the professional to leverage their experience and knowledge.
- Ability to use the engine and the intelligence to search multiple content sets that might be stored by numerous parties in various databases that are searched using numerous search engine technologies.

What it's not

- Replace Boolean searching – instead it couples Boolean with state of the art machine learning techniques to create a unique, powerful and most importantly scaleable solution.
- Disintermediate the professional searcher or their experience/expertise – instead it offers a platform that casts a more comprehensive net from which searchers can work quickly to find the documents that are most relevant.

Technology is rarely the solution by itself. The best role of technology is to empower the user in ways that make them more effective.


Semantic Search - Windows Internet Explorer
http://www.lexisnexis.com/totalpatent/searchSemantic.do?semantic='y'
File Edit View Favorites Tools Help
Share Browser WebEx
Semantic Search
Home
TotalPatent™
NEW Enhancements to TotalPatent!
Project ID: test client
Search
Document Retrieval
History & Alerts
Analytics
Work Folders
Results
Guided Search
Advanced Search
Semantic Search
Notes Search
Enter at least 3 search items below, with no Boolean connectors. You can enter these items as words, phrases, or sentences. For best results, enter homogeneous terms that relate to one concept at a time. For example, "mechanical heart valve". Tell me more ...
Search Input
Enter or paste text here. Do not use connectors. Limit of 32,000 characters.
Search Within
Full Text (incl. Biblio.)
latex athletic neoprene
Select Preview
Preview Results
Search Now
Reset form
Search Options
Also search for terms in English machine translations
Remove family member duplicates
Check Settings
Publication Number Search
Enter a List
Upload a List
Enter 1-500 Publication Numbers
View accepted publication number formats
View
Results list
Search
Look Up Assignee or Inventor
Search for variations of assignee or invento names, then add them to your patent search.
Assignee
Inventor
Find
Find Subsidiary Companies

Semantic Search - Windows Internet Explorer

http://www.lexisnexis.com/totalpatent/searchSemanticSubmit.do

File Edit View Favorites Tools Help

Share Browser WebEx

Semantic Search | Home

TotalPatent™

Powered by PureD

Preview Semantic Terms and Results

Change and re-analyze your search terms, click or drag a semantic term to change its status in your query, and see how your changes affect your results. Learn More

Your Search Terms

latex athletic neoprene

Regenerate Terms

Terms Generated by Semantic Analysis

Undo **20** of **20** terms used Add another te

REQUIRED (Boolean "AND")

latex

OPTIONAL (Boolean "OR")

athletic **neoprene**

footwear rubatex athletic shoes
neoprene foam rubber soles cush
insole wearer's foot outsole
overshoes leather running shoes
vamp

EXCLUDED (Boolean "NOT")

HOLDING AREA (Not included in your search)

Most relevant **20** of **9330** results: Retrieve All Results

US20110277349A1 2011-11-17 UNIBODY CONSTRUCTION
FOOTWEAR AND METHOD FOR MAKING THE SAME

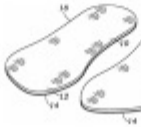
sh Abstract:
resent application discloses footwear
rising a body structure in which at least upper
de of one continuous folded composite
rial comprised of layered sheets of material.

EP1502517B1 2012-01-11 Articulated welt footwear constru
and related method of manufacture

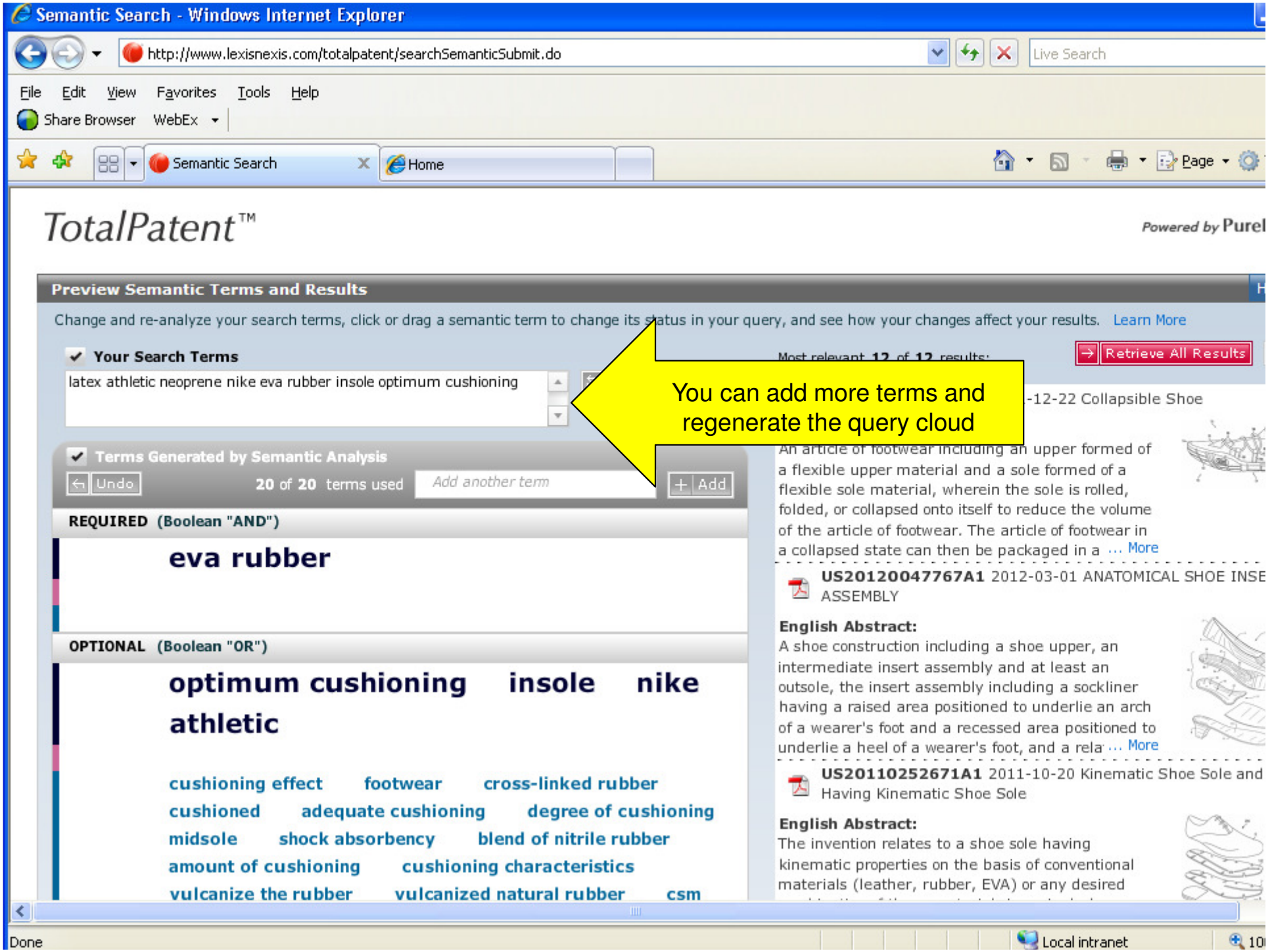
sh Abstract:
wear construction including an articulated welt. The footwear
es an upper and an outsole secured together with a welt in th
rd pertion of the footwear. The welt terminates short of the he
of the footwear. In one embodiment, the outsole in the heel
is seeured to a heel cradle which is further secured the uppe
otwear provides the durability of a welt construction in the ···

US20110283562A1 2011-11-24 INSOLE FOR FOOTWEAR

sh Abstract:
resent disclosures concerns embodiments of a
ear insole that can be used with various types
of footwear, including, without limitation, shoes
(including open and closed toe shoes), boots,
sandals, etc. The insole includes an upper fabric
layer that comes in contact with the foot and ... More



Query Cloud (on the left)
Returns words or phrases "inferred" by the Semantic Brain The size and color of the words indicate the weighting of that word in the search and Previews top 20 Results (on the right) if these terms were run without modification


Semantic Search - Windows Internet Explorer
http://www.lexisnexis.com/totalpatent/searchSemanticSubmit.do
Live Search
File Edit View Favorites Tools Help
Share Browser WebEx
Semantic Search
Home
Page
TotalPatent™
Powered by Pure
Preview Semantic Terms and Results
Change and re-analyze your search terms, click or drag a semantic term to change its status in your query, and see how your changes affect your results. Learn More
Your Search Terms
latex athletic neoprene nike eva rubber insole optimum cushioning
Retrieve All Results
You can add more terms and regenerate the query cloud
Terms Generated by Semantic Analysis
Undo
20 of 20 terms used
Add another term
Add
REQUIRED (Boolean "AND")
eva rubber
OPTIONAL (Boolean "OR")
optimum cushioning insole nike athletic
cushioning effect footwear cross-linked rubber
cushioned adequate cushioning degree of cushioning
midsole shock absorbency blend of nitrile rubber
amount of cushioning cushioning characteristics
vulcanize the rubber vulcanized natural rubber csm
-12-22 Collapsible Shoe
An article of footwear including an upper formed of a flexible upper material and a sole formed of a flexible sole material, wherein the sole is rolled, folded, or collapsed onto itself to reduce the volume of the article of footwear. The article of footwear in a collapsed state can then be packaged in a ... More
US20120047767A1 2012-03-01 ANATOMICAL SHOE INSE ASSEMBLY
English Abstract:
A shoe construction including a shoe upper, an intermediate insert assembly and at least an outsole, the insert assembly including a sockliner having a raised area positioned to underlie an arch of a wearer's foot and a recessed area positioned to underlie a heel of a wearer's foot, and a rela ... More
US20110252671A1 2011-10-20 Kinematic Shoe Sole and Having Kinematic Shoe Sole
English Abstract:
The invention relates to a shoe sole having kinematic properties on the basis of conventional materials (leather, rubber, EVA) or any desired
Done
Local intranet

Search Results - Windows Internet Explorer

http://www.lexisnexis.com/totalpatent/resultsPage.do?layout=resultsSingle

File Edit View Favorites Tools Help

Share Browser WebEx

Search Results | Home

TotalPatent™

Project ID: test client2 | Sign Out | Preferences | Contact Us | Hel

Search | Document Retrieval | History & Alerts | Analytics | Work Folders | Results

Search Terms REQUIRED: ("eva rubber"[H]) OPTIONAL: "athletic"[H], "insole"[H], "nike"[H ...)

View Search Query | Edit Search | Save Search | Create A

View Analyze | Purchase | Citation Map

Layout

You can also run a search within your results

Narrow Search GO

Using Semantic Concepts What's this?

esults: **12** of 58,374,330 searched

1 to 12

mp to Documents 1 to 12 GO

How Do I

All

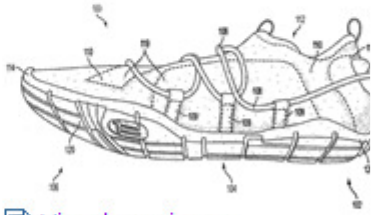
Show Term Hits: **Off** | On

Fields | Sort

1 **US20110308107A1** 2011-12-22 **Collapsible Shoe**

Application Number:

English Abstract:
An article of **footwear** including an upper formed of a flexible upper material and a sole formed of a flexible sole material, wherein the sole is rolled, folded, or collapsed onto itself to reduce the volume of the article of **footwear**. The article of **footwear** in a collapsed state can then be packaged in a container. This container can be dispensed by a vending machine in a convenient urban area.

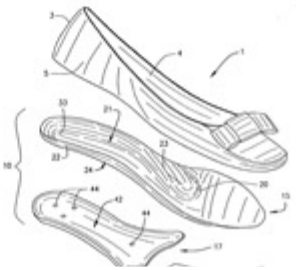


View large image

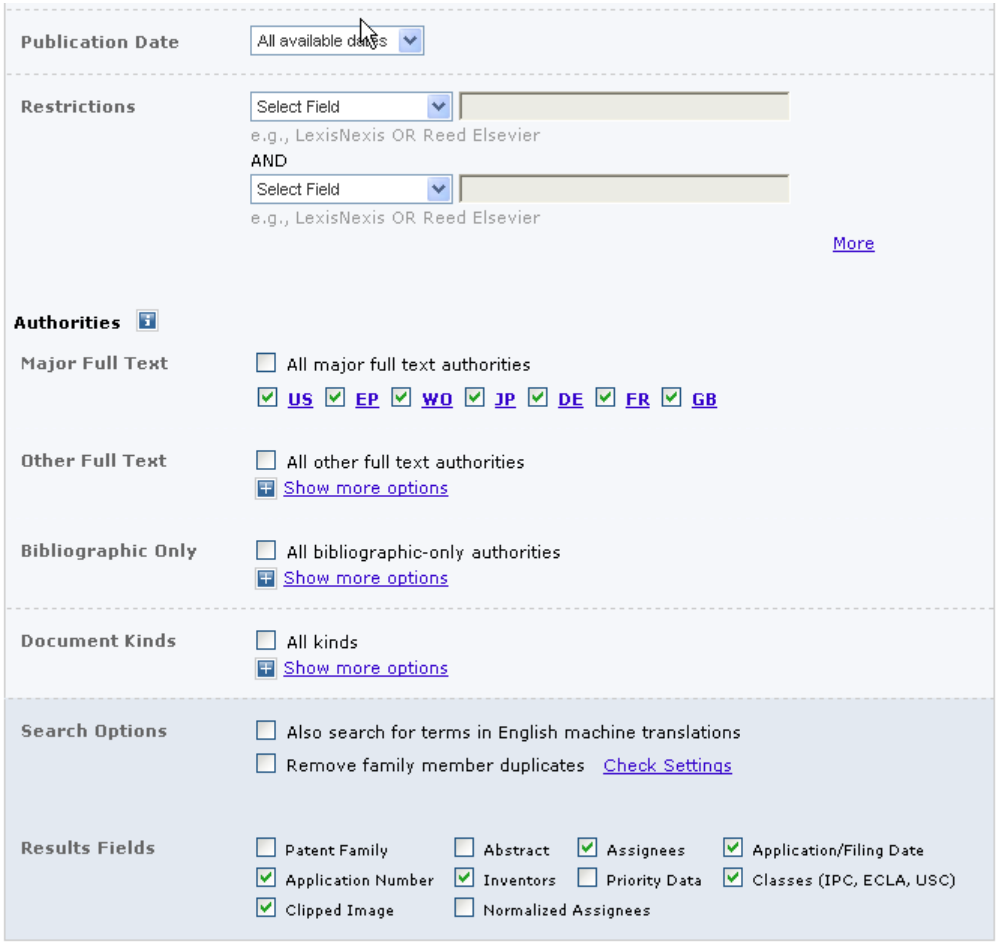
2 **US20120047767A1** 2012-03-01 **ANATOMICAL SHOE INSERT ASSEMBLY**

Application Number:

English Abstract:
A shoe construction including a shoe upper, an intermediate insert assembly and at least an outsole, the insert assembly including a sockliner having a raised area positioned to underlie an arch of a wearer's foot and a recessed area positioned to underlie a heel of a wearer's foot, and a relatively rigid board member having recessed areas underlying the sockliner in the heel and arch areas.



Restriction Options



Bottom half of the Restrictions screen allows the searcher to further restrict and narrow the query same as in the Advanced Search Form

Review of Semantic Search Tools for Patent Research

- Moradei and Contessini (Associazione Italiana Documentalisti Brevettuali) – EPO Patent Information Conference 2011, Kilkenny Ireland

1. Relevant documents overlooked
2. Different Search Strategies return different documents
3. Retrieval based upon general search terms and terms are weighted equally
4. Keyword strategies produce noisy results lists

LexisNexis® Responses

- Relevant documents overlooked –
  - Role of semantic search is ***not to replace Boolean*** but to ***supplement*** it
    - ***expanding*** the results pool to find documents that may be relevant but not found with standard Boolean searching
  - Documents that are related due to common cites, inventors or assignees may not share common concepts. Semantic search focuses upon terms and phrases related to ***core concepts***, not upon matching documents.
  - Ranking of semantic searches is imperfect and selection of the top few documents returned may overlook that patent which may be closely related but in a different domain.

LexisNexis® Responses

- Different Search Strategies return different documents
  - Goal is to give the user control over their search. Search strategies are an important component in this control.
  - Term weighting, term selection and exclusion all contribute to the user's ability to expand the range of their searches.
  - Restrictions can focus search as they do with Boolean searches but can also lose possible "discoveries".

LexisNexis® Responses

- Keyword strategies produce noisy results lists
  - Current research is considering elimination of the mandatory use of “required” terms to allow the concepts to define the search domain.
  - Example: a search on “pizza box” with a required phrase of “pizza box” does a good job of identifying food containers and not network computers BUT
  - Without the mandatory phrase, it finds “**cylindrical packing containers**” which leads us to:
  - **EP 2 062 824 B1 Folded box for transport of cylindrical articles**

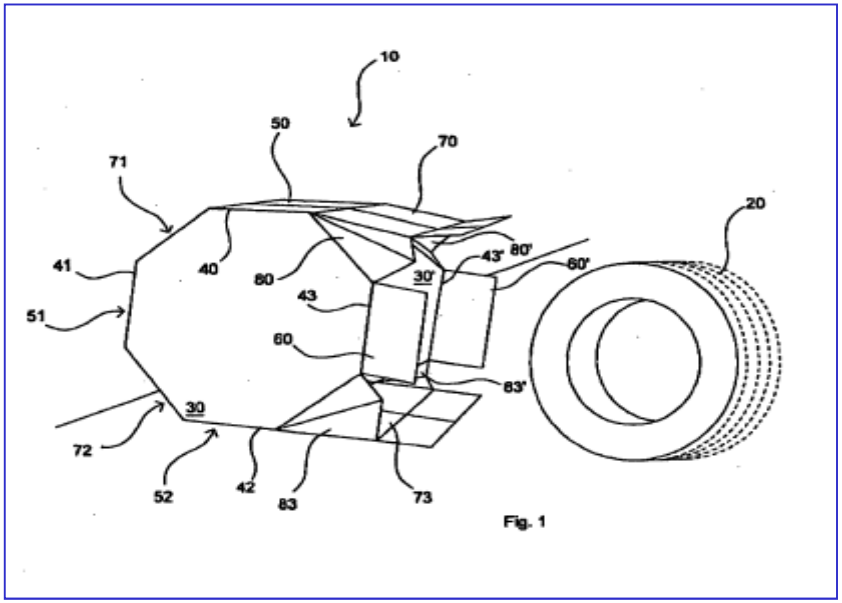


Fig. 1

LexisNexis® Responses

- Retrieval based upon general search terms and terms are weighted equally
    - Moving to a new search platform has altered the effectiveness of term weighting. Recent and upcoming releases will re-emphasize the differences between terms selected as “high”, “medium” and “low”