

A nighttime photograph of a city street, likely in Ottawa, Canada. The scene features a row of historic buildings with many lit windows. In the foreground, a street with light trails from moving vehicles is visible. A tall clock tower is prominent in the background on the right. The overall atmosphere is urban and illuminated by streetlights and building lights.

November 22nd, 2014

Semantics for SharePoint

An Introduction to Semantics for Sharepoint

Luca Scagliarini
CEO Expert System USA
Expert System
lsagliarini@expertsystem.com

SHAREPOINT SATURDAY
OTTAWA

Semantics for SharePoint

An Introduction to Semantics for Sharepoint

Luca Scagliarini
CEO Expert System USA
Expert System
lscagliarini@expertsystem.com

SHAREPOINT SATURDAY
OTTAWA

What We'll Cover

Maximize “findability” in SharePoint

Contrast and Compare Methods

- Context
- What is the Advantage
- Show it working (if you are still here)

Thought for the Day

Search is an activity, and find is a result.

Which do you think the users prefer?

What are we talking about?

Delivering This

The business wants this



The users tag like this



Search returns this

Search returns this



What's the problem?

- Users don't tag
- If they don't tag SharePoint is not as effective delivering content.
- Makes search unproductive.
- Per several research made "38% of work is searching for information"
38% is a big number if you calculate the cost of their time

How much i\$ that?

Consider the possible savings of reducing time searching for enterprise content by just 30 minutes each week. Based on 1,000 associates working a US average of 1,700 hours per year over 50 weeks, you will have contributed the productive equivalent of nearly 15 additional employees.

To quantify that, let's assume that the average loaded cost of an employee is \$60,000 per annum. In dollar terms, you will have contributed the equivalent of more than \$882,000 in the first year. If \$35,000 is the average cost, then the contribution is more than \$514,000, if \$80,000, your contributed value is \$1,176,000.

Easy Fix

The answer is...

Please see that your seat back and tray table are in their upright and locked position...

Keyword search is like a game show



How Its Been Done

The Keyword

Az Expert System szokásosan nagy hangsúlyt helyez a továbbképzésre, így munkatársai évente számos szakmai tanfolyamon vesznek részt. Az elmúlt években a csoport több tagja is részt vett több hónapos, egyesült államokbeli, angliai illetve németországi projekt munkákban, melyek során nemzetközi csoportban végeztek fejlesztői tevékenységet

Does it matter if it's Hungarian?
They're still words Right?

How Its Been Done

Shallow Linguistics

Words in text are recognized as either belonging to the dictionary, or not.

- Acknowledged words are associated to a basic headword and a grammatical type is assigned.
- Limited logical groupings can be made.
- Indexes contain headwords and defined keywords.

Or

If I have it listed, I can probably work with it.

These Technologies Force a Compromise “Four door fuel economy”



Precision (a measure of exactness)
Retrieving precise but limited results
(this meets your criteria everything else was excluded)

Recall (a measure of completeness)
Retrieving a large number of potentially
Relevant results but not very precise (here is everything I
found, what you are looking is probably in here somewhere)

Limitations of a Technology Driven Approach

Technology does not understand the meaning of words.

Keyword Technology
or Statistics



- **Breaks text into single words without considering the context**, like reading a language that you don't understand:

• *Az Expert System szokásosan nagy hangsúlyt helyez a továbbképzésre, így munkatársai évente számos szakmai tanfolyamon vesznek részt.*

Systems understand languages like java, C#, COBOL, basic, etc.
Not English or Hungarian!

Shallow Linguistic
Technology



Recognizes words and identifies their most basic forms (lemmas), but cannot distinguish between different meanings.

Sell -> Selling -> Sold

Meaning = context, sentiment, and relevance.

The Word of the Day is Disambiguation

Morphological Analysis	understand word forms	dog, dog-catcher and doggy-bag are closely related
Grammatical Analysis	understand the parts of speech	"There are 40 rows in the table" uses rows as a noun, vs. "She rows 5 times a week" uses rows as a verb.
Logical Analysis	how words relate to other words	Davey Jones, represented by attorney Daniel Stanley, is married to Rebecca Carter". Rebecca is married to Davey, not Daniel.
Semantic Analysis (disambiguation)	understand the context of words	"I used chicken broth for my soup stock" uses stock in the context of food, vs. "The company keeps lots of stock on hand" uses stock in the context of inventory.

Semantic Analysis and Disambiguation

- For people, the meaning of a word is understood by its definition and the surrounding elements further refine the understanding of context (the way the word is used).

Disambiguation The Game

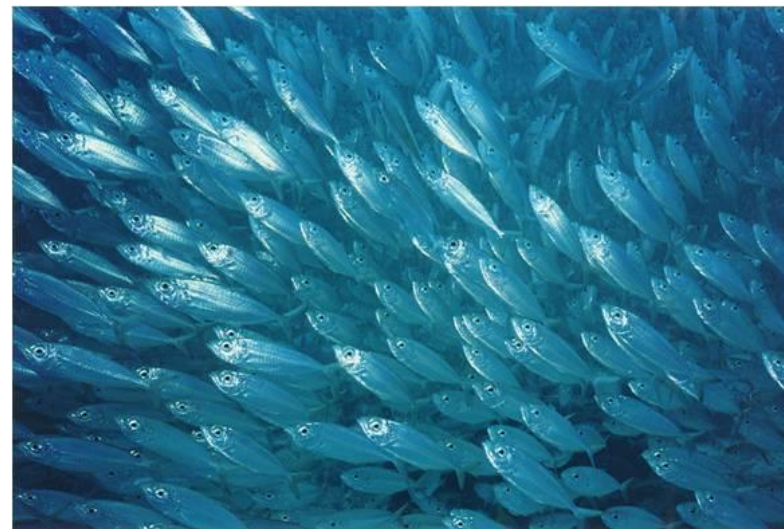


Let's Play
Guess The Meaning

Contestant Number One

- Swimming in a school.
- Quick, What's the meaning?

That's a trick question Luca!
The real question is what's
the context?



It's Not Definition, It's Context!

Why Is Meaning Important?

- Meaning in Context = High Quality Automated Content Tagging
- High Quality Automated Content Tagging = Accurate and Granular Automated Classification
- Accurate and Granular Automated Classification = Findable Content

- Findable Content =



Not Easy to Understand, Not So Easy for a User to Tag

National saving rates and domestic investment rates. The correlation between the two, for countries representing four-fifths of world GDP, declined from a coefficient of around 0.96 in 1992, where it had hovered for a half-century, to an estimated low of 0.8 last year.¹ A number of factors have recently converged to lessen restraints on cross-border financial flows as well as on trade in goods and services.

You should not have to be a subject
matter expert to find content

You should be able to find the subject matter
expert from the content

Tagging Is More Than Frequency

- Tagging requires an understanding of how the content relates to organization (node(s))
- Tagging also requires an understanding of associated relationships (taxonomy)
- Best practices =
 - Consistent tagging optimized for taxonomy

Semantics can relieve the burden of tagging from the user, empower the administrator, and provide faceted content navigation as augmented search

What does semantics do for me

<actions><act syn="#77495" lemma="need" phrase="2" kind="v. of state"><clause>
This article needs additional citations for <span
class="s27590">verification</clause><l role="other" syn="#27590"
lemma="verification" kind="cstm:Event"/></act><act syn="#69431"
lemma="challenge" phrase="4" kind="v. of social interaction"><clause> <span
class="v308411">Unsourced material may be <span
class="s69431">challenged [...] </clause><l role="obj" vsyn="#78448"
syn="#v308411" lemma="Unsourced" kind="concept"/></act><act syn="#66737"
lemma="remove" phrase="4" kind="v. of action"><clause> [...] and <span
class="s66737">removed</clause><l role="obj" vsyn="#78448"
syn="#v308411" lemma="Unsourced" kind="concept"/></act><act syn="#77408"
lemma="be" phrase="22" kind="v. of state"><clause> <span
class="v308413">SharePoint is a web
application framework and platform [...] </clause><l role="sbj" vsyn="#37940"
syn="#v308413" lemma="Microsoft SharePoint" kind="organization"/></act>

I Thought We Were Talking About SharePoint

The screenshot displays the 'Cogito Enriched Document Library' interface. The top navigation bar includes 'Site Actions', 'Browse', 'Documents', and 'Library'. The breadcrumb trail shows 'Cogito Connected for SharePoint > Cogito Enriched Document Library > All Documents'. A search bar on the right says 'Search this site...'. The left sidebar contains links for 'Cogito Connected', 'Libraries', 'Lists', 'Discussions', 'Recycle Bin', and 'All Site Content'. The main content area shows a table of documents with the following columns: Type, Name, Document Abstract, People, Organizations, Places, Standard Categories, IPTC Categories, Last Analysis Status, and Locked. A filter 'Last Analysis : 9/26/2014 (4)' is applied. Four documents are listed:

Type	Name	Document Abstract	People	Organizations	Places	Standard Categories	IPTC Categories	Last Analysis Status	Locked
Document	4.6.15_19.xml	Report: Extremist groups busted in northwest China; (CNN) -- Police in northwest China's Xinjiang Uyghur Autonomous Region have busted 23 "terror and religious extremism groups and caught over 200 suspects", state news agency Xinhua reported, citing the re		Cable News Network; Xinhua	China; Xinjiang; Urumqi; Kashgar; Hotan; Aksu	Social science; Anthropology; Ethnology; Politics; Social services; Crime; Police	Unrest, Conflicts and War; Civil Unrest; Social Conflict; Act of Terror; Economy, Business and Finance; Media; News Agency; Counter Terrorism; Politics; Defence; National Security	No	
Document	4.6.14_17.xml	Mesmerizing video shows star explode;Hubble's 4-year exploding star timelapse					Arts, Culture and Entertainment; Cinema; Disaster and Accident; Emergency Incident; Explosion	No	
Document	4.6.14_18.xml	Spain fans' agony at bitter defeat;Spain thrashing brings agony to Madrid			Spain; Madrid			No	
Document	4.6.14_19.xml	Still, the ruling Scottish National Party believes that Scotland should be separate; that its foundations are not only sturdy enough to support an independent state, but	Campbell Tartan; David Cameron; Jose Manuel Barroso; Margaret Thatcher; Robert Burns	CNN International; North Atlantic Treaty Organization; United Nations; Bank of England	Germany; United Kingdom; England; London; Northern Ireland; Scotland; Edinburgh; Wales; Spain; United States of America; City of Westminster	Politics; Economics; Institutes	Economy, Business and Finance; Economy (general); Macro Economics; Exports; Politics; Diplomacy	No	

The Windows taskbar at the bottom shows the time as 8:03 PM on 10/7/2014.

We're Talking About Supercharging SharePoint with Semantics

The screenshot shows a SharePoint Search Center interface with the following components:

- Browser Bar:** Address bar shows `95.110.200.168/Search/results.aspx?k=politics`.
- Site Actions:** Includes 'Browse' and 'Page' tabs.
- Enhanced Search:** Search bar contains 'politics'. A 'Preferences' link is available.
- Results Summary:** '1-10 of 18 results'.
- Standard Categories:**
 - ☒ Any Standard Categories
 - Politics (13)
 - Law (5)
 - Social services (4)
 - Military (3)
 - Art (2)
 - Social science (2)
 - Economics (1)
 - Geography (1)
 - Telecommunications (1)
 - Transports (1)
- Apply:** Button to filter results.
- Search Results:**
 - Israeli military suspects teens missing from Jewish settlements kidnapped**
Turner Broadcasting System Inc.Cable News Network ... Social servicesCriminal lawLawPoliticsTelecommunicationsBroadcasting eventsMathematics ...
Authors: Sdadmin34684242\Admin34684242 Date: 6/14/2014
<http://95.110.200.168/Lists/Team News/DispForm.aspx?ID=5>
 - Ukrainian official: Russian tanks have entered the country**
Turner Broadcasting System Inc.Foreign OfficeCable News NetworkBritish Broadcasting CorporationFight ... MilitaryPoliticsWorkTelecommunicationsTransportsBroadcasting eventsMinistriesPublic ...
Authors: Sdadmin34684242\Admin34684242 Date: 9/18/2014
<http://95.110.200.168/Lists/Team News/DispForm.aspx?ID=4>
 - World Cup: Netherlands thump Spain in Salvador**
Turner Broadcasting System Inc.Spanish TV stationXabi AlonsoCable News NetworkBarcelonaBayern MunichChelsea ... SportFootballGameWorkBroadcasting eventsPoliticsEthnologyShowGeneticsLaw ...
Authors: Sdadmin34684242\Admin34684242 Date: 9/18/2014
<http://95.110.200.168/Lists/Team News/DispForm.aspx?ID=3>
 - Obama says no combat troops to Iraq: U.S. weighs airstrikes**
Turner Broadcasting System Inc.Washington Institute for Near East PolicyCentral Intelligence ... PoliticsMilitaryAir forceWorkEconomicsSportMedicineMinistriesIslamFinanceCrimeSociology ...
Authors: Sdadmin34684242\Admin34684242 Date: 7/1/2014
<http://95.110.200.168/Lists/Team News/DispForm.aspx?ID=2>
 - 1815787.txt**
Tehran Radio Lets Critics Vent Over Iran's Nuclear ... Arts, Culture and EntertainmentRadioPoliticsNuclear PolicyReligion and BeliefReligious LeaderPoliticsGovernmentNational Government ...
Authors: Sdadmin34684242\Admin34684242 Date: 6/16/2014 Size: 6KB
<http://95.110.200.168/Enriched Document Library/1815787.txt>
 - 4.6.14_19.xml.txt**
- People:**
 - ☒ Any People
 - George W. Bush (2)
 - Isis (2)
 - Saddam Hussein (2)
 - Abdullah El-Shamy (1)
 - Abu Musab Al-Zarqawi (1)
 - Abu Ubaida (1)
 - Ahmed Aboul Gheit (1)

The right sidebar features a word cloud with terms like 'United States of America', 'China', 'Iraq', 'Beijing', 'Syria', 'Iran', 'Germany', 'Baghdad', 'London', 'Egypt', 'New York', and 'Hong Kong'.

That Concludes The Entertainment Portion of the Program

- If you like
- You may ask any question I can answer
- See the semantic technology working live (No sales pitch) I promise
- And accept my thanks for participating!

And now a word from our sponsors

Thank you to all of our Sponsors!!



O'REILLY®

WILEY



Remember to fill out your evaluation forms to win some great prizes!

&

Join us for SharePint today!

Date & Time: Immediately after prizes
Location: Observatory Student Pub
Address: On Algonquin Campus
Parking: No need to move your car!*

Site: <http://www.algonquinsa.com/ob.aspx>

Thank you

Luca Scagliarini

lscagliarini@expertsystem.com

@scagliarini