



Collaborative Knowledge Discovery & Marshalling

for Intelligence & Security Applications

Russell Jensen

Andrew Cowell, Michelle Gregory, Peter Ellis, Alex Franks, Eric Marshall, Keith Fligg,
Liam McGrath, Kelly O'Hara, Kelly Domico, Ryan Hull, Eric Bell

May 26th 2010

IEEE Workshop on Current Issues in Predictive
Approaches to Intelligence and Security Analytics



Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

Information Resources

Modern Publishing Corporation
 Department of Computer Games Technology
 House 2009, Acton, CT 06810, 10 pages
 800-233-0000/06810

Research Article Combining AI Methods for Learning Bots in a Real-Time Strategy Game

Robin Baumgarten, Simon Colton, and Mark Helm

Department of Computing, Imperial College London, London SW7 2BZ, United Kingdom
 {r.baumgarten, s.colton, m.helm}@imperial.ac.uk

Correspondence should be addressed to Robin Baumgarten, robin.baumgarten@imperial.ac.uk

Received: 18 Mar 2009, Accepted: 21 October 2009

Revised: 18 Mar 2009, Accepted: 21 October 2009

home about index RC Wiki archive start here Search

RealClimate

Climate science from climate scientists

20 Top Blog WINGS Science

2007 Science & Technology WEB AWARDS

RealClimate Investment Network

Technical Note: We have just upgraded the blog software. It looks ok, but please let us know



Modell and Sir



Country	Year	Value 1	Value 2	Value 3	Value 4	Value 5	Value 6	Value 7	Value 8
Canada	2006	5355541	12454747	0.43	1	1	0	1	2
Canada	2004	206020	7007482	0.0294	1	1	0	1	2
Canada	2005	210736	7167902	0.0294	1	1	0	1	2
Canada	2006	215399	7326496	0.0294	1	1	0	1	2
Dominican	2004	683229	7007482	0.0975	1	1	-99	-99	2
Dominican	2005	698870	7167902	0.0975	1	1	-99	-99	2
Dominican	2006	714333	7326496	0.0975	1	1	-99	-99	2
Mexico	2004	25829	6593527	0.0388	0	0	-99	-99	0
Mexico	2005	260413	6711676	0.0388	0	0	-99	-99	0

twitter Home Profile Find People Settings Help Sign out

What are you doing? 140

Latest: @bafanks have fun being all dolled up! 1 day ago

Home

- @Entourage1 Entourage! http://bit.ly/Sib1A about 1 hour ago from web
- @NewPost Post: Race for Life- support a good cause http://tinyurl.com/3urdj about 1 hour ago from twitterfeed
- @Gh-tri-cities wind storms, how I don't miss you when I'm away. about 1 hour ago from twitterfeed
- @sucks I'm on my way to Long Beach and can't beg a piece from you. about 2 hours ago from twitterfeed
- @http://twitpic.com/66h55 "Single-girl-at-a-wedding" Re: Raspberries, Blackberries, and my grandma's in-pan pie shell. about 3 hours ago from Twitter
- @Ryan thinks this Butterfinger ice cream is good for his (mental) health... thinks this Butterfinger ice cream is good for his (mental) he... about 7 hours ago from twitterfeed
- @InCase Power Slider vs Mophie Juice Pack A [GIVEAWAY INSIDE]! http://tinyurl.com/ig07nm about 7 hours ago from twitterfeed
- @Listening to Jackson 5 and baking pie. about 8 hours ago from web
- @Off to see "Rod Blagojevich Superstar" downtown. Good times will be had! http://tinyurl.com/deq9ae about 10 hours ago from web

Forum Table:

Forum	Last Post	Threads	Posts	Moderator
Indian Real Estate Talk	09-25-2007 06:13 PM	5	22	manish, manish, Cactus
Adobe Creative.Suit	Today 12:55 PM	565	2,089	manish, manish, Cactus
Need to promote this Forum	07-04-2009 01:10 PM	39	261	manish, Cactus, manish
Rel.Power.IPO.sensex.an.Jan...	01-03-2008 06:31 AM	126	603	manish, Cactus
Environmental.Hazardous.Sub...	Today 09:03 AM	85	277	manish, Cactus
How.We.Increase.The.Speed.Off...	07-18-2009 06:22 PM	55	262	manish
Multi-Architecture.Semantic.Tax...	Today 05:39 AM	125	259	manish
Real.Estate.Writing.Archieve	Today 06:11 AM	192	511	manish
How.can.we.know.which.will.b...	Yesterday 03:50 PM	157	359	manish
Unfilled.seats.offered.Fora...	Today 06:06 AM	45	152	manish, Cactus, manish
Powerful.System.For.Indians...	Today 10:38 AM	370	846	manish, Cactus
The.most.suitable.job.fel...				

Mission

- ▶ Allow modelers to **exploit** the ‘information resources’ they have and **discover** new, potentially relevant material across new media types
- ▶ Knowledge Encapsulation Framework (KEF) aims to provide:
 - an effective method for **storing**, retrieving, reviewing and annotating your documents
 - an environment where you can **share** these materials with team members and discuss
 - a mechanism to **discover** new, related information from social and traditional media
 - a means to **link** this material to model representations to aid analysis and game-play
- ▶ We do this via a **semantic wiki**, enabled with an **NLP pipeline**, other components and user participation



Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

The KEF Semantic Wiki

- ▶ A wiki is a web site whose pages can be edited by authorized users via a browser
 - Many wiki 'engines' – we use Mediawiki, arguably the most popular wiki engine, used by Wikipedia and Intellipedia
 - We utilize the semantic mediawiki extension, SMW+ and a host of other community developed resources, plus some homegrown



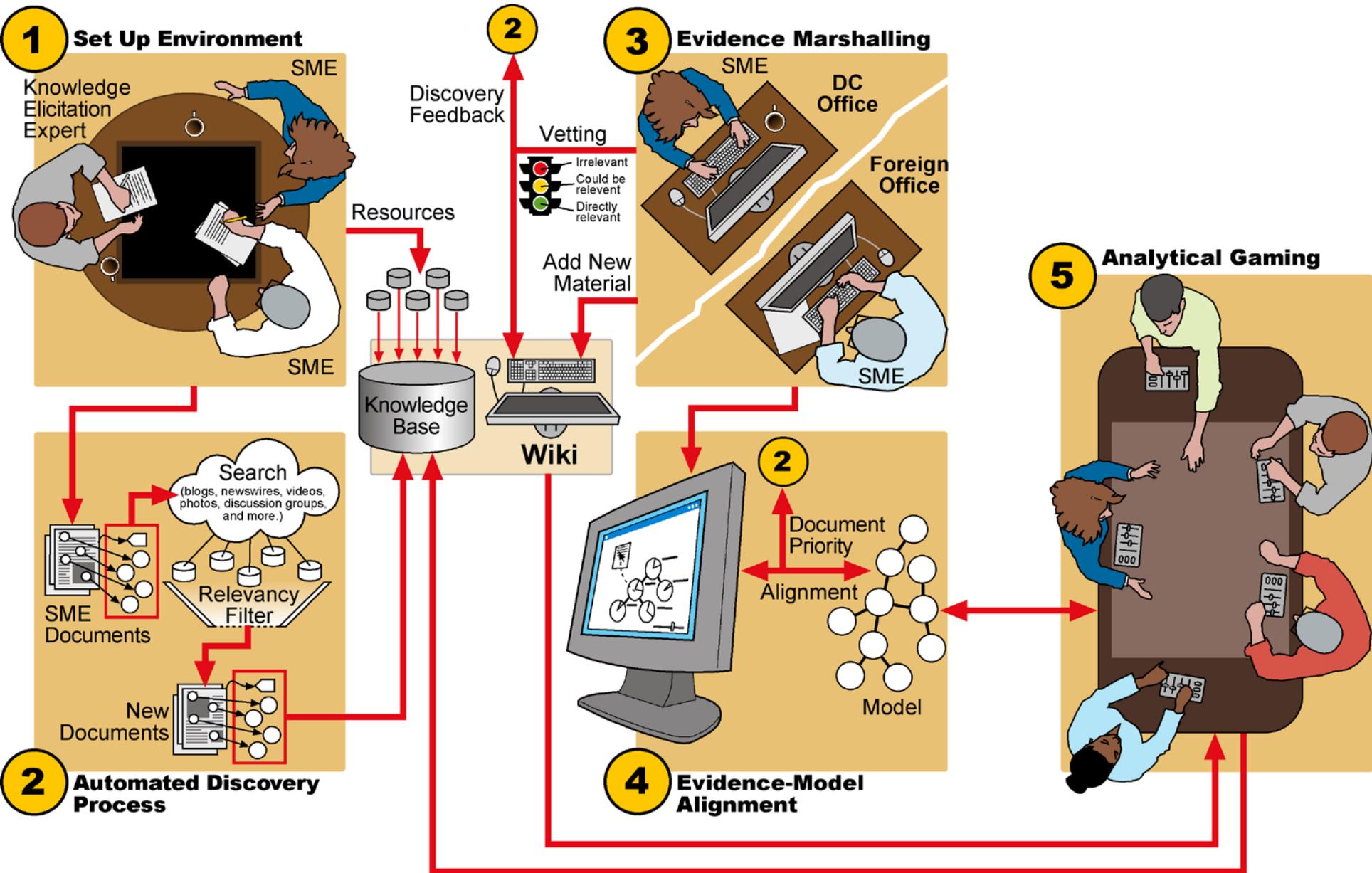
Pacific Northwest
NATIONAL LABORATORY

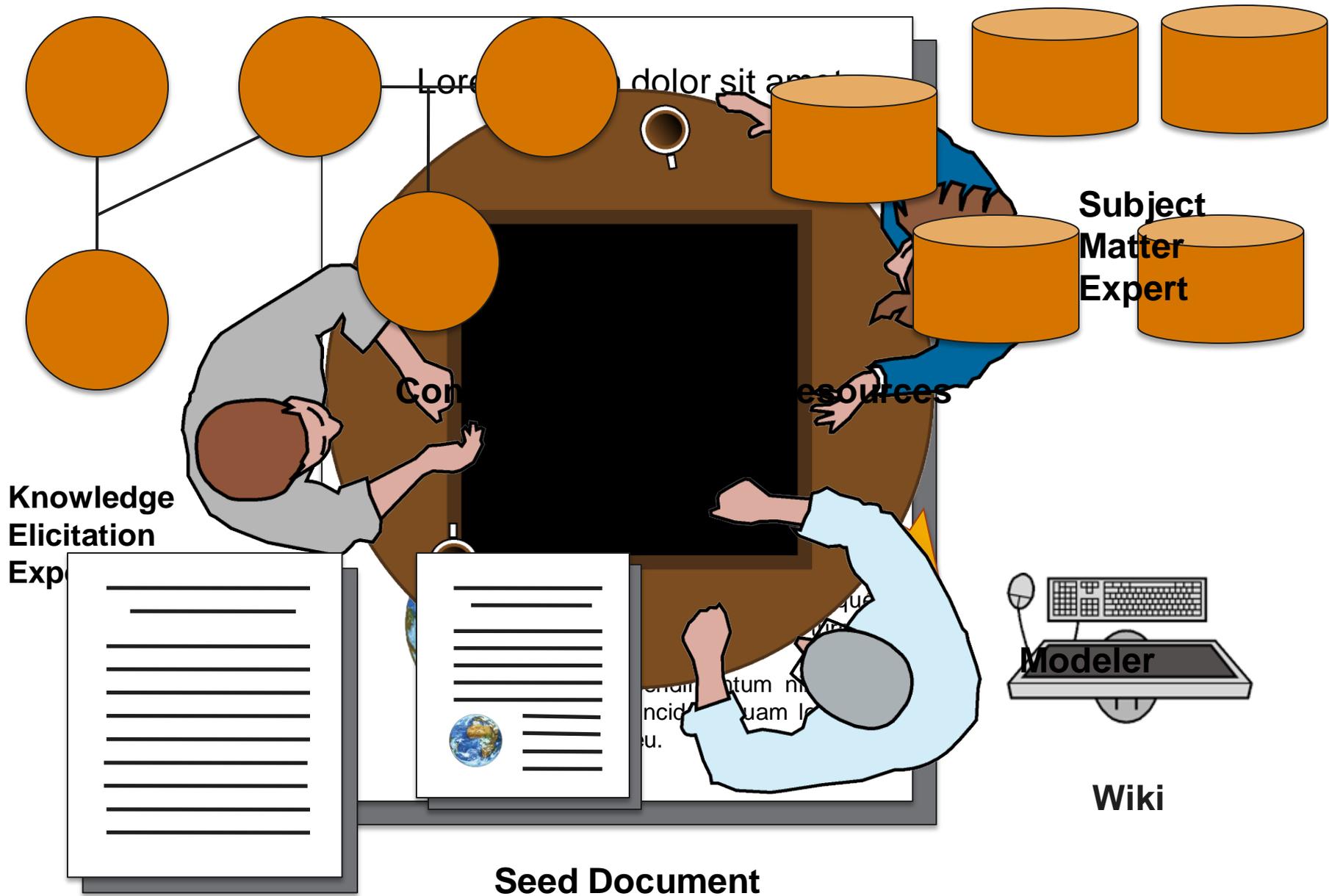
Proudly Operated by Battelle Since 1965

The KEF Semantic Wiki

- ▶ Semantic Wiki's extend this definition by enabling us to define relationships within the text:
 - 'The population of Richland is 38708' becomes:
 - 'The population of Richland is [[population::38708]]'
- ▶ We're adding structure to the unstructured to enable both humans and machines to understand the content
- ▶ Utilizing combined machine and human annotation to achieve this goal







2. SME Discovery Process

Title

Lorem ipsum dolor sit amet

Jane Smith John Doe

Author

Topics

Pellentesque interdum facilisis diam, vel posuere tellus dignissim id. In viverra cursus erat, eget malesuada arcu **accumsan** tristique. Sed lacinia rhoncus justo. Mauris tincidunt tincidunt sem vitae **pellentesque**.

Keywords

People and Groups

Nam **posuere**, risus feugiat scelerisque vulputate, est diam sollicitudin diam, vitae **laoreet dui** diam vel felis. Nam **neque risus**, imperdiet ut feugiat id, **pellentesque** nec urna.

Places



Maecenas porta, orci vel feugiat lobortis, mi odio lacinia magna, sed pellentesque mauris arcu non turpis. Aenean luctus condimentum nibh, eu tincidunt quam lobortis eu.

Events

Seed Document



Search Term

Title

Lorem ipsum dolor sit amet

Jane Smith John Doe

Author

Topics

Pellentesque interdum facilisis diam, vel posuere tellus dignissim id. In viverra cursus erat, eget malesuada arcu **accumsan** tristique. Sed lacinia rhoncus justo. Mauris tincidunt tincidunt sem vitae **pellentesque**.

Keywords

People and Groups

Nam **posuere**, risus feugiat scelerisque vulputate, est diam sollicitudin diam, vitae **laoreet dui** diam vel felis. Nam **neque risus**, imperdiet ut feugiat id, **pellentesque** nec urna.

Places



Maecenas porta, orci vel feugiat lobortis, mi odio lacinia magna, sed pellentesque mauris arcu non turpis. Aenean luctus condimentum nibh, eu tincidunt quam lobortis eu.

Events

Seed Document



Search Term



IEEE

SCIENCE

ISI WEB OF KNOWLEDGESM

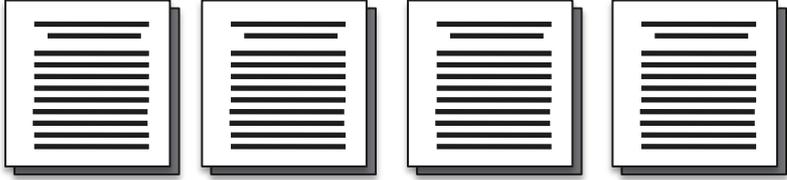


Social Media

2. Discovery Process

Traditional Media

Technical Reports | Journals | Conference Articles



Social Media

Blog | Photo | Video | Forum



Relevancy Filter

Lorem ipsum dolor sit amet

Jane Smith John Doe

Pellentesque interdum facilisis diam, vel posuere tellus dignissim id. In viverra cursus erat, eget malesuada arcu accumsan tristique. Sed lacinia rhoncus justo. Mauris tincidunt tincidunt sem vitae pellentesque.

Nam posuere, risus feugiat scelerisque vulputate, est diam sollicitudin diam, vitae laoreet dui diam vel felis. Nam neque risus, imperdiet ut feugiat id, pellentesque nec urna.



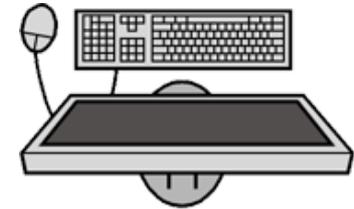
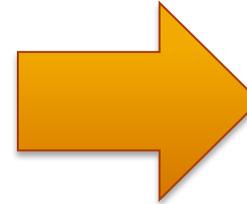
Maecenas porta, orci vel feugiat lobortis, mi odio lacinia magna, sed pellentesque mauris arcu non turpis. Aenean luctus condimentum nibh, eu tincidunt quam lobortis eu.

Properties

Title
Author
Keywords
Topics
People
Groups
Places
Events
Rhetoric
Sentiment

+

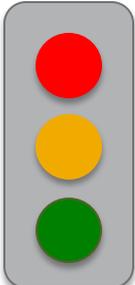
Full Text



Wiki

-  Irrelevant
-  Could Be Relevant
-  Directly Relevant

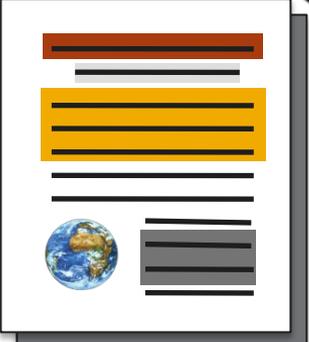
Joint Global
Change Research
Institute (MD)



Add New Material



Pacific Northwest
National Laboratory
(WA)



3. Review Wiki

Challenge

- ▶ We have a set of evidence that support a node in a model
- ▶ We also have a corpus of documents that are related
- ▶ Can we use the evidence we already have to automatically find additional evidence in the corpus that support the same idea?



Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

Current Approach

- ▶ Develop or implement an algorithm that compares two sentences semantically to determine if they are closely related.
- ▶ Compare statement of evidence with other sentences in the corpus to find sentences that support the same idea.



Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

Sentence Similarity

- ▶ Tag the all sentences to be compared with parts of speech
- ▶ Perform a word sense disambiguation on all the words
- ▶ Change word forms to those found in WordNet
- ▶ Drop all stop words and words not found in WordNet
- ▶ Compute sentence similarity by doing a semantic comparison using WordNet, weighted by word frequency in a large corpus, combined with word order similarity between sentences
- ▶ The sentences are ranked according to sentence similarity.

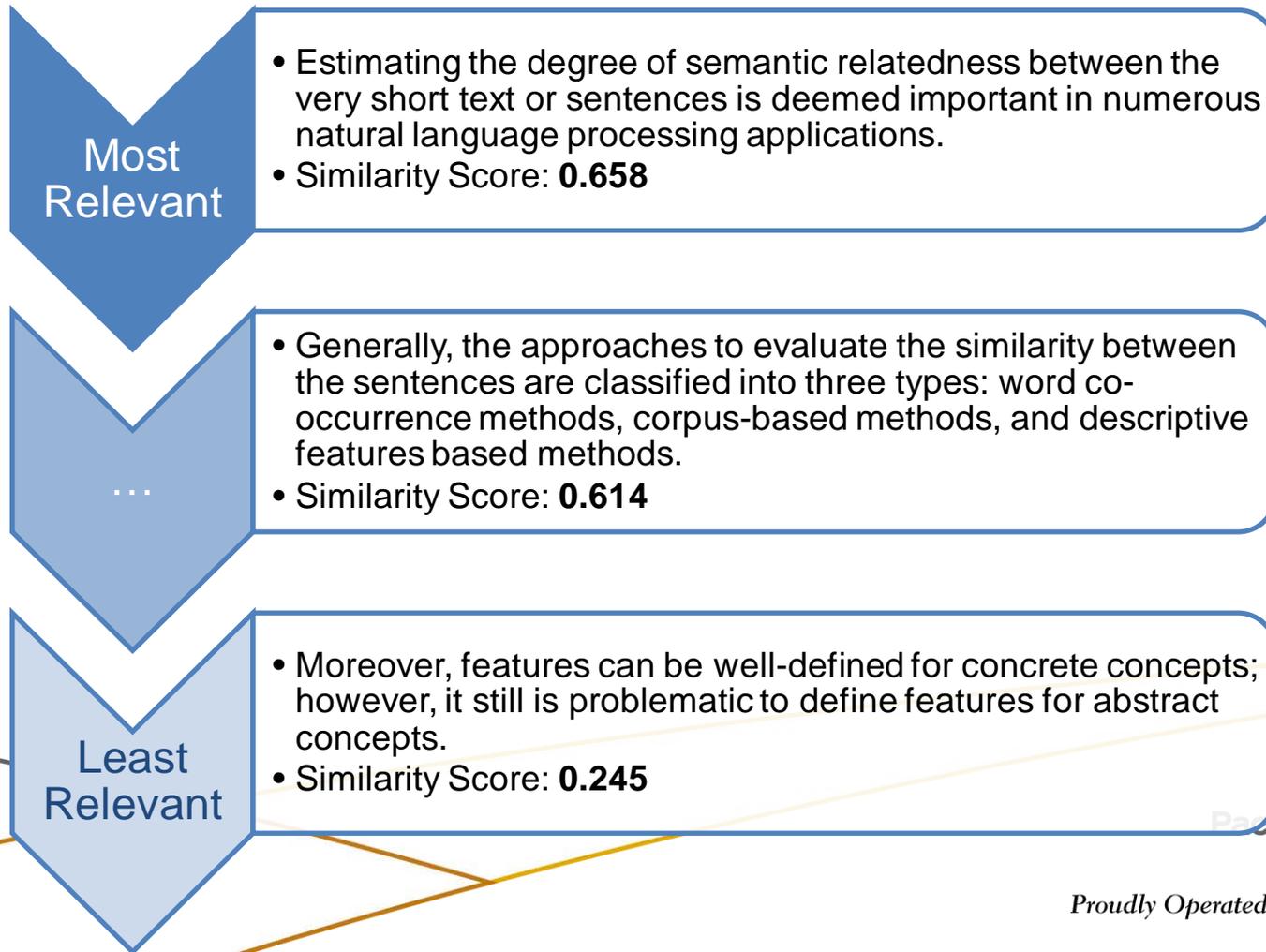


Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

Results

- ▶ Seed: Recent applications of natural language processing present a need for an effective method to compute the similarity between very short texts or sentences.



Results

- ▶ Seed: The threat that terrorists could acquire and use a nuclear weapon in a major U.S. city is real and urgent.

Most Relevant

- A dangerous gap remains between the urgency of the threat of nuclear terrorism and the scope and pace of the U.S. and world response.
- Similarity Score: **0.551**

...

- In Russia and the other states of the former Soviet Union, there is some good news to report, but there is still far too much bad news.
- Similarity Score: **0.547**

Least Relevant

- The focus on HEU should not lead the world community to ignore the burgeoning global stockpiles of separated civilian plutonium.
- Similarity Score: **0**

Access to Education

Sed lacinia rhoncus justo. Mauris tincidunt tincidunt vitae pellentesque.

Maecenas porta, orci vel feugiat lobortis, mi odio lacinia magna, sed pellentesque

Priority:

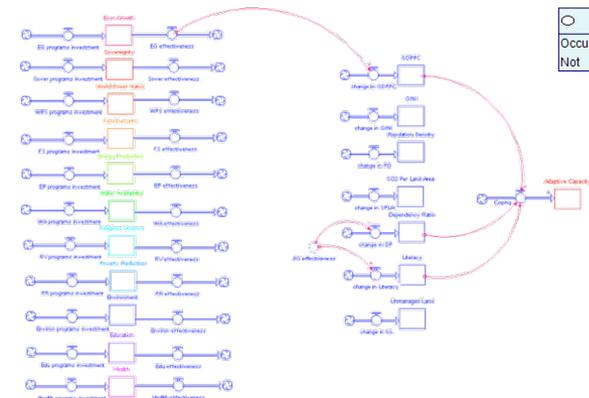
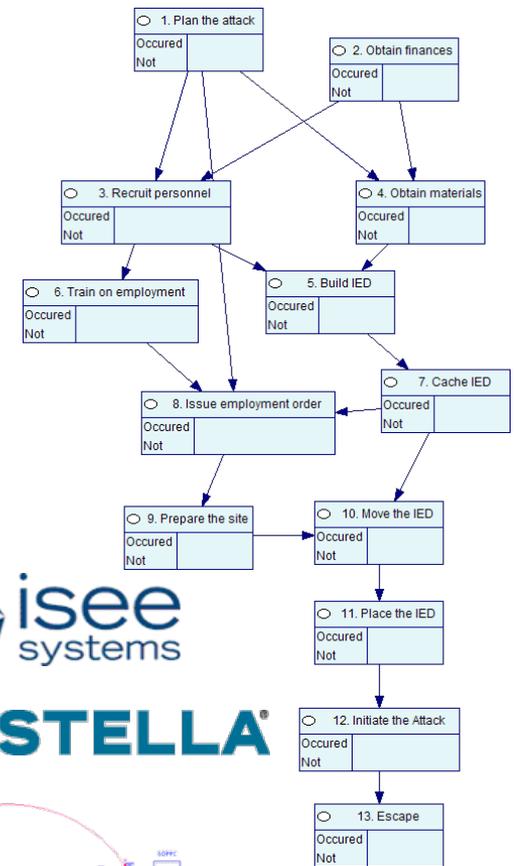
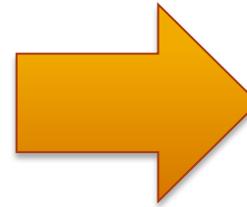


Potentially Related Evidence

Nam neque risus, imperdiet ut feugiat id, pellentesque nec urna



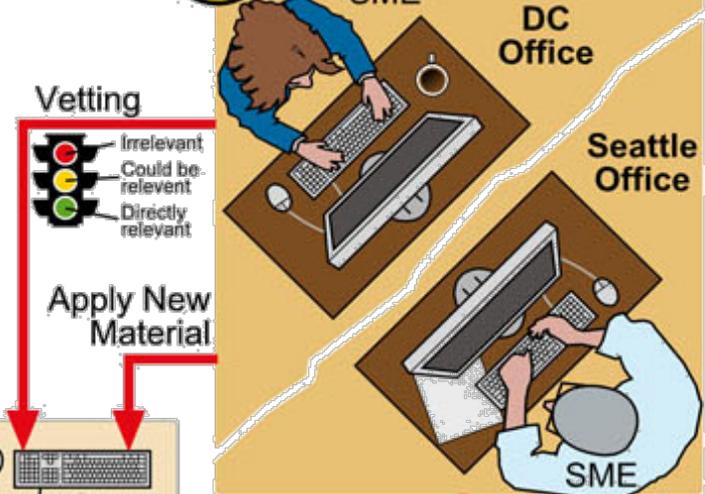
Pellentesque interdum facilisis diam, vel posuere tellus dignissim id.



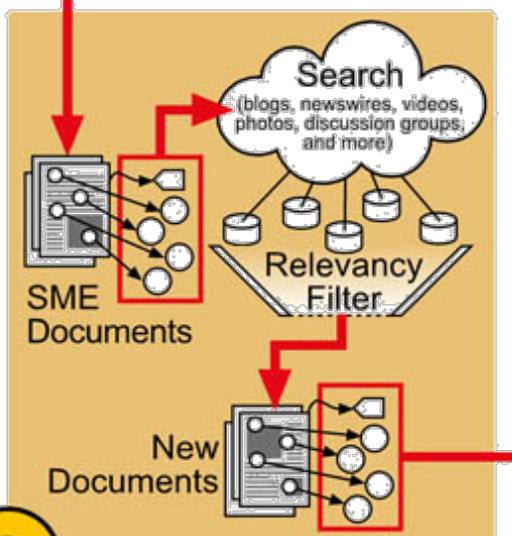
1 Subject Matter Expert (SME) Discussion



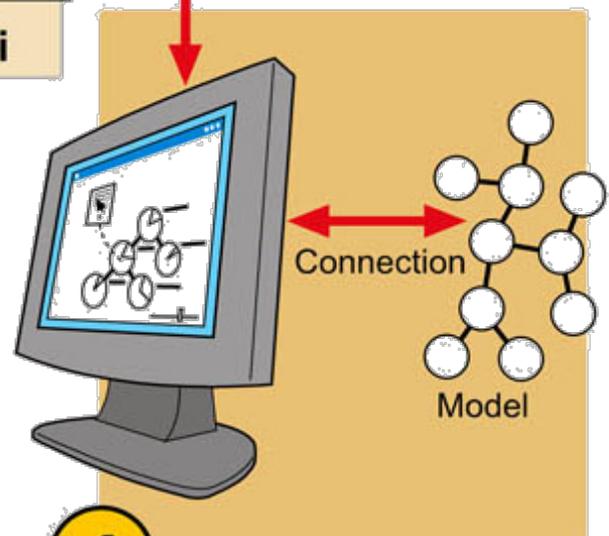
3 Evidence-Marshalling Environment



2 Automated Discovery Process



4 Evidence-Model Alignment



Novelty of Approach

- ▶ We've developed a collaborative, integrated evidence marshalling environment for modeling & simulation
- ▶ Novel in that it:
 - combines social and traditional media together in a unified framework for analysis
 - integrates an automated discovery mechanism to find new, potentially related material
 - leverages machine learning techniques to help prioritize incoming information resources
 - uses a combined machine/user approach to tagging information resources, extensible based on domain of application
 - utilizes cutting edge semantic web technologies
 - “Probably less than 100 [public] semantic wiki’s in the world” (Koren, 2009)
 - provides context-sensitive visualization mechanisms

