



# Mainstream Semantic Enrichment

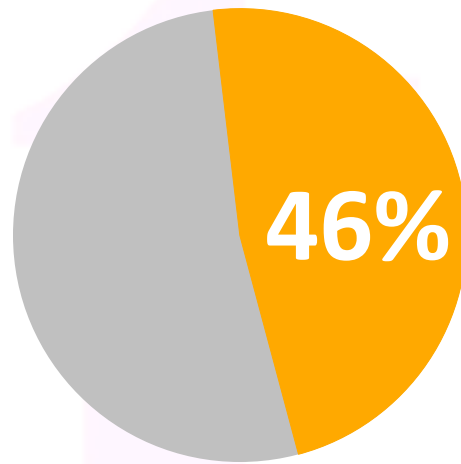
A key platform for your growth  
now easier than ever before

Daniel Mayer,  
VP Corporate Marketing

INTERNATIONAL ASSOCIATION OF SCIENTIFIC, TECHNICAL & MEDICAL PUBLISHERS  
The voice of research publishing since 1969

*stm*

# Strong Adoption

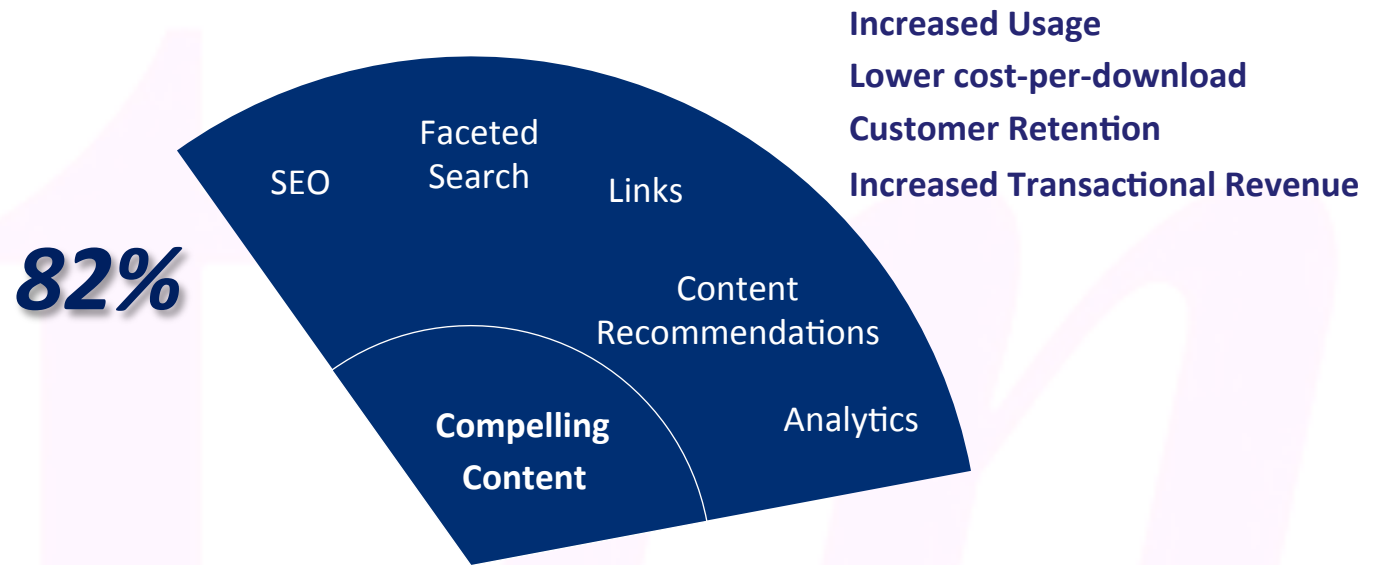


semantically enrich  
their content

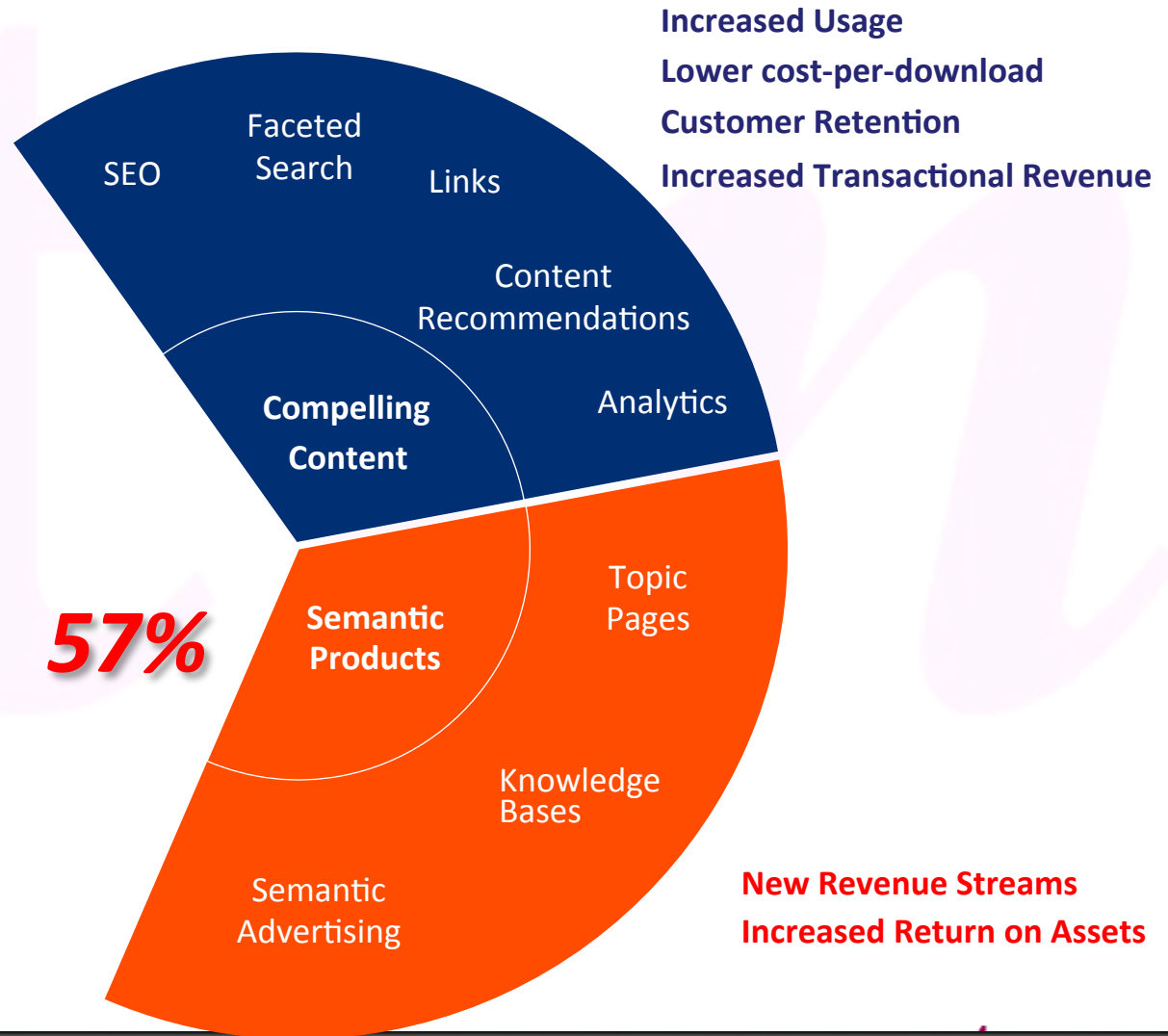
Publishers  
of Scientific Journals

*Source : 2011 study by the Publishing Research Consortium*

# Where is the value today ?

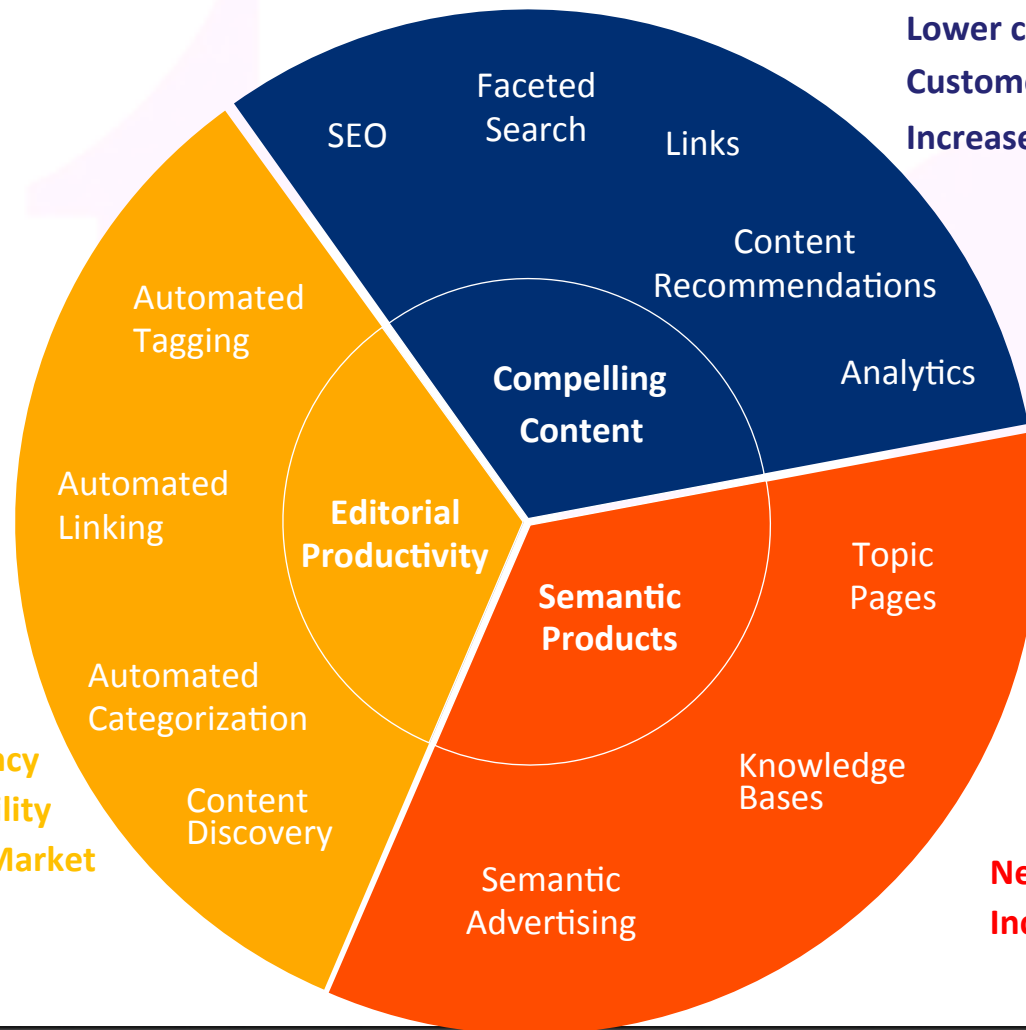


# Where is the value today ?



# Where is the value today ?

Increased Usage  
Lower cost-per-download  
Customer Retention  
Increased Transactional Revenue



# Future value

Providing analytical insight

Serving authors

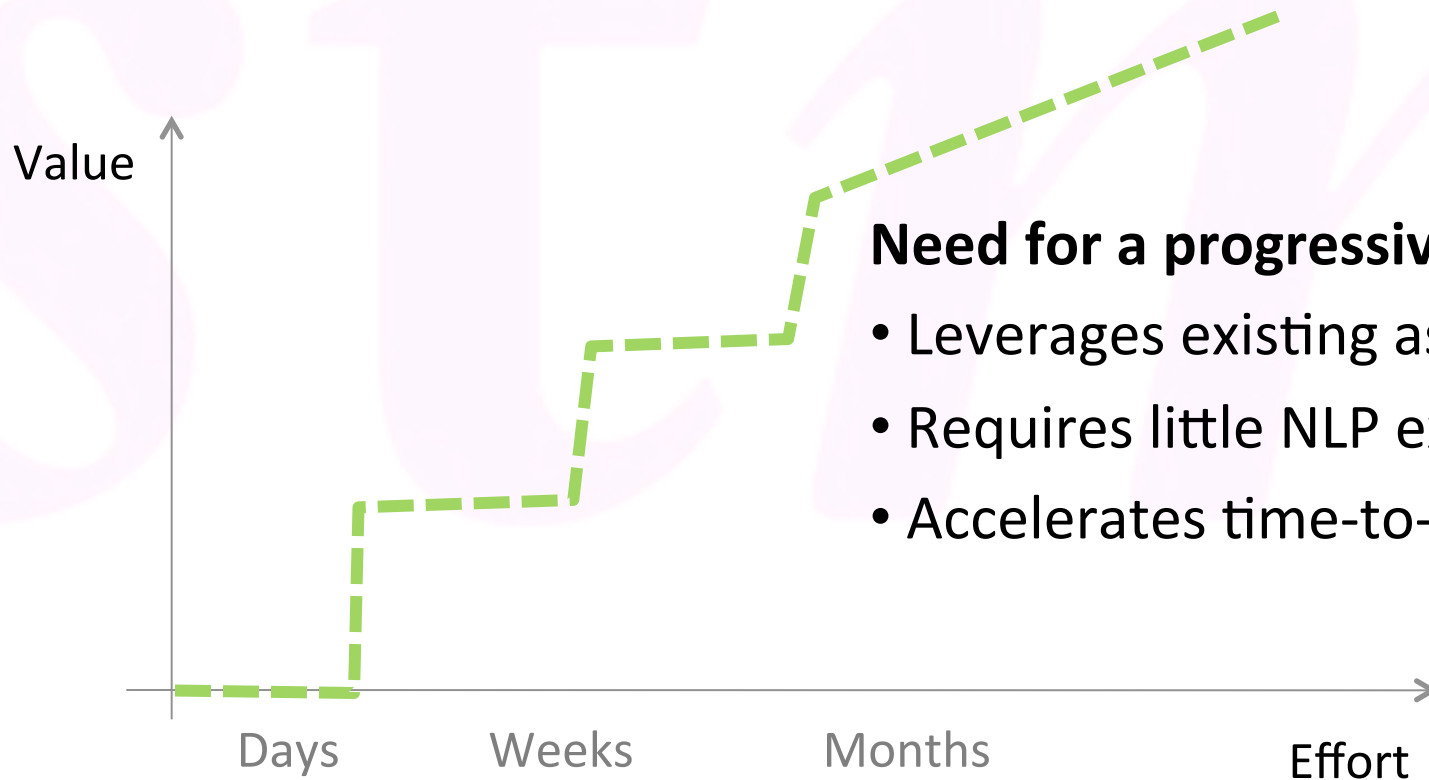
Weaving proprietary & commercial content together

Delivering in the workflow

# Where's the challenge ?



# The path to mainstream



## Need for a progressive ROI

- Leverages existing assets
- Requires little NLP expertise
- Accelerates time-to-value



# Smart Taxonomy Facilitation

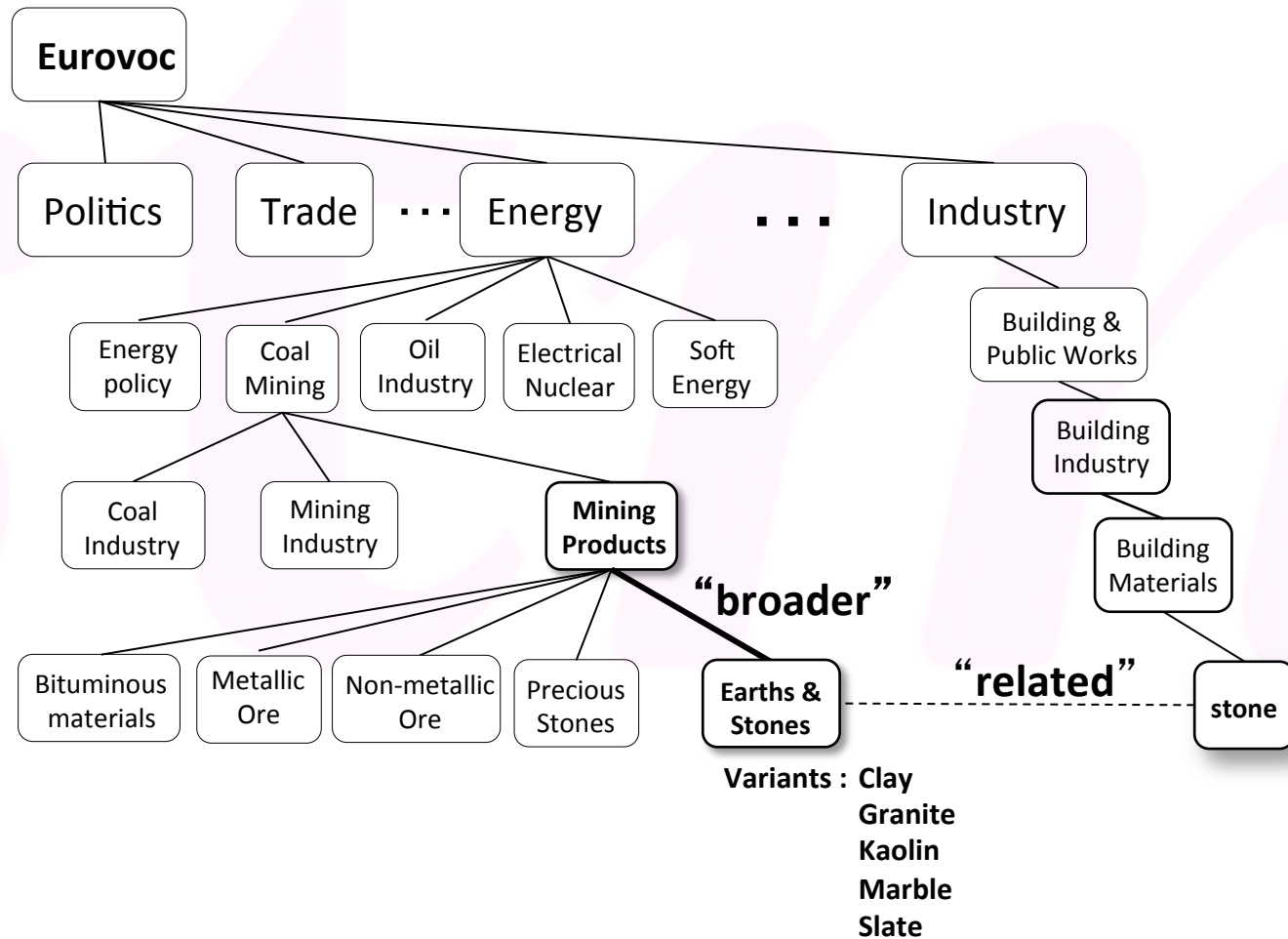
Leverages existing taxonomy/ontology/thesaurus

Circumvents weaknesses

Requires no NLP expertise

Build an extractor in a matter of hours

# Taxonomy : an inside look



# Taxonomy **weaknesses**

## **Silence**

- False negatives caused by lack of variants

## **Noise**

- False positives caused by ambiguous terms
- Example : « orange »

**Three techniques** help circumvent the weaknesses of naive string matching

# 1. Fuzzy Term Matching vs. Silence

Identify / Score variants of taxonomical concepts

		score
<b>Preferred form</b>	Acousto-optical deflector	1.0000
<b>Permutation</b>	<b>Optico-acoustic</b> deflector	0.8677
<b>Insertion</b>	Acousto-optic <b>based</b> deflector	0.8339
<b>Spelling alternatives</b>	Acousto-optical deflector	0.9259

Benefit : **Increased Recall**

## 2. Relevance Scoring vs. Noise

Rank the relevance of each extracted concept based on

- **Position** in the document or in a particular zone
- **Frequency** in the document
- **Relative Frequency** (idf) in a reference corpus
- **Node degree** - a measure of connectedness of the term with related taxonomical concepts also found in the document
- **Length** - number of tokens of the term
- **Depth** of the concept in the taxonomy
- **Reliability** of the term (customizable)

Benefit : **Increased precision**

# 3. Part of Speech Tagging vs. Noise

Case Type	Example	Potential issue	Luxid® solution
<b>Longest match</b>	ice cream	/Ice /Cream	/Ice Cream Chunking process identifies noun phrases and longest match heuristic retains the longer forms contained in these noun phrases as the most relevant extraction.
<b>Defactorization</b>	Cow and sheep farming	/Sheep Farming	/Cow Farming /Sheep Farming Part-of-speech tagging layer recognizes and expands factorized forms. Analysis layer extracts all parts of the factorized forms.
<b>Part-of-Speech Ambiguity</b> (adjective vs. noun)	Sole proprietor	/Fish/Sole	Part-of-speech tagging layer recognizes the word sole is an adjective and not a noun. Analysis layer rejects the match with the fish name.
<b>Part-of-Speech Ambiguity</b> (verb vs. noun)	can lead	/Container/Can /Metal/Lead	Part-of-speech tagging layer recognizes these words are verbs not nouns. Analysis layer <i>rejects the matches</i> .

Benefit : **Increased precision**

# Importing a taxonomy

Knowledge Editor 6.0.0 190

File Window Knowledge Base Help

Knowledge Entities Extraction

Filter on /Entity/EVTerm starting with  Ok

Page 1 Test Extraction

1096 entities

	broader	label	related	variant
1049	/Entity/EVTerm/1050	inshore grounds		
1342		installation allowance		
1341	/Entity/EVTerm/165	insurance indemnity	/Entity/EVTerm/1339	
1192		interest group	/Entity/EVTerm/1197	lobby;pressure group
1436		interference		duty to interfere;right to interfere
1034		International Atomic Energy ...	/Entity/EVTerm/1455	IAEA
1024		International Energy Agency		IEA
1016		international finance	/Entity/EVTerm/1194	
1286		International Institute for Edu...		IIEP
1043	/Entity/EVTerm/1455	International Monetary Fund		IMF
1044		International Social Security ...		ISSA
11		international trade	/Entity/EVTerm/1807	world trade
1132		investment protection		investment guarantee
1295	/Entity/EVTerm/1184	Ionian Islands		Corfu
1417		iron and steel industry		electrical steelworks;foundry;iron and steel undertaking;iron and steel works;stee...
1287		island		
1457		judicial investigation		confidentiality of judicial investigations
1443		labour inspectorate		factory inspectorate;industrial inspectorate
1178		large business		large undertaking
1179		large holding		
1388		leather industry		leather production;tanning industry
108		leave on social grounds	/Entity/EVTerm/106	
1138	/Entity/EVTerm/1282	leftism		
1080	/Entity/EVTerm/1538	legal expenses		
1438		legislative initiative		
1407		light industry		
1205		linguistic group		
1017	/Entity/EVTerm/1018	local authority finances	/Entity/EVTerm/1317	rench-speaking areas;Frenc...
1317	/Entity/EVTerm/1310	local tax	/Entity/EVTerm/1017	nces
1161	/Entity/EVTerm/1154	logistics		taxation
1002	/Entity/EVTerm/1000	long-term financing		
1371	/Entity/EVTerm/1410	machine-tool industry	/Entity/EVTerm/1724	
1154		management	/Entity/EVTerm/1070;115	
1155		management accounting	/Entity/EVTerm/1164	
1160		management of resources	/Entity/EVTerm/1159	

Add new entity :  +

Edit values:

Value		
local taxation	-	
regional tax	-	
regional taxation	-	

Cartridges

- STF
- Thesaurus
- Static Mapping

# Importing a taxonomy

Knowledge Editor 6.0.0 190

File Window Knowledge Base Help

Knowledge

- Demo
  - Configuration (0)
  - Thesaurus (0)
  - Entity (0)
    - language
    - variant
    - EVTerm (1096)
      - broader
      - Domain
      - label
      - label.french
      - label.german
      - related
      - Thesaurus
      - ThesaurusID
      - variant.french
      - variant.german

Cartridges

- STF
  - Thesaurus
  - Static Mapping

Entities Extraction

Filter on /Entity/EVTerm starting with  Ok

Page 1 Test Extraction

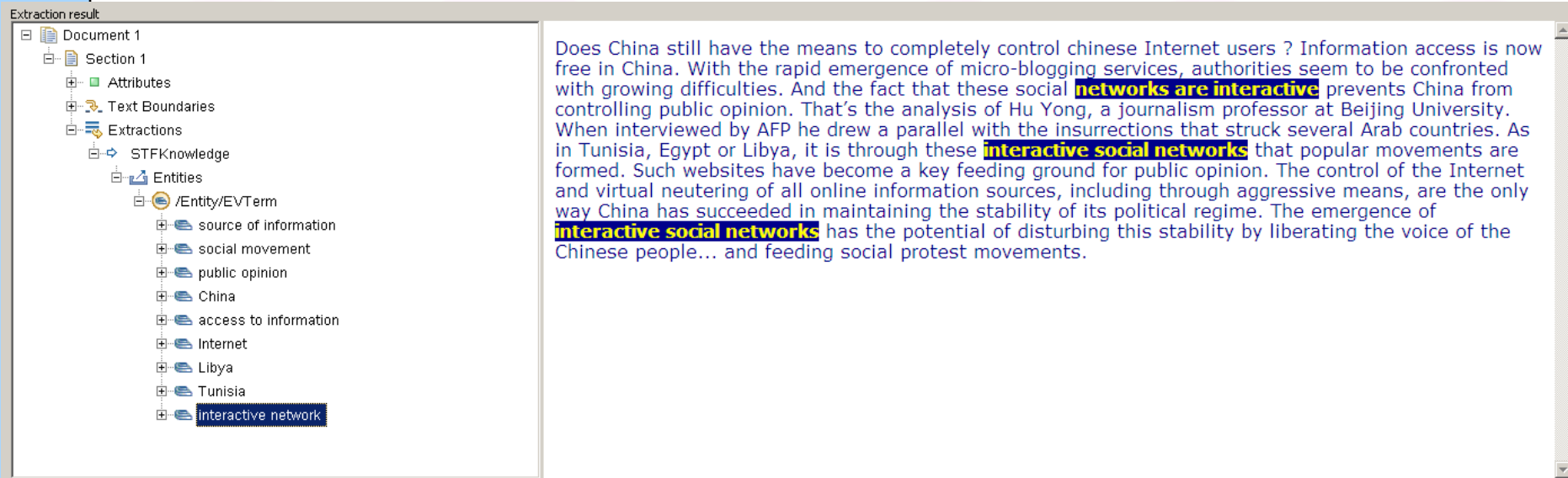
1096 entities

	label	label.french	label.german
1049	inshore grounds	fond côtier	küstennaher Meeresboden
1342	installation allowance	indemnité d'installation	Einrichtungsbeihilfe
1341	insurance indemnity	indemnité d'assurance	Versicherungsleistung
1192	interest group	groupe d'intérêt	Interessengruppe
1436	interference	ingérence	Einnischung
1034	International Atomic Energy Agency	Agence internationale de l'énergie atomique	Internationale Atomenergie-Organisation
1024	International Energy Agency	Agence internationale de l'énergie	Internationale Energie-Agentur
1016	international finance	finances internationales	internationale Finanzen
1286	International Institute for Educational Plan...	Institut international de planification de l'éducation	Internationales Institut für Bildungsplanung
1043	International Monetary Fund	Fonds monétaire international	Internationaler Währungsfonds
1044	International Social Security Association	Association internationale de la sécurité sociale	Internationale Vereinigung für soziale Sicherheit
11	international trade	commerce international	internationaler Handel
1132	investment protection	garantie des investissements	Investitionsschutz
1295	Ionian Islands	Îles ioniennes	Ionische Inseln
1417	iron and steel industry	industrie sidérurgique	Eisen- und Stahlindustrie
1287	island	île	Insel
1457	judicial investigation	instruction judiciaire	Verfahren vor dem Untersuchungsrichter
1443	labour inspectorate	inspection du travail	Gewerbeaufsicht
1178	large business	grande entreprise	Großunternehmen
1179	large holding	grande exploitation	landwirtschaftlicher Großbetrieb
1388	leather industry	industrie du cuir	Lederindustrie
108	leave on social grounds	congé social	Urlaub aus sozialen Gründen
1138	leftism	gauchisme	Linksradikalismus
1080	legal expenses	frais de justice	Gerichtskosten
1438	legislative initiative	initiative législative	Gesetzesinitiative
1407	light industry	industrie légère	Leichtindustrie
1205	linguistic group	groupe linguistique	sprachliche Gruppe
1017	local authority finances	finances locales	Gemeindefinanzen
1317	local tax	impôt local	örtliche Steuer
1161	logistics	gestion du matériel	Materialverwaltung
1002	long-term financing	financement à long terme	langfristige Finanzierung
1371	machine-tool industry	industrie de la machine-outil	Werkzeugmaschinenindustrie
1154	management	gestion	Management
1155	management accounting	gestion comptable	betriebliches Rechnungswesen

Add new entity :  +



# Fuzzy Term Matching example



The screenshot displays a software interface for document analysis. On the left, a tree view under 'Extraction result' shows a document structure: 'Document 1' contains 'Section 1', which has 'Attributes', 'Text Boundaries', and 'Extractions'. Under 'Extractions', there is 'STFKnowledge', which contains 'Entities'. The 'Entities' list includes: '/Entity/EVTerm', 'source of information', 'social movement', 'public opinion', 'China', 'access to information', 'Internet', 'Libya', 'Tunisia', and 'interactive network'. The 'interactive network' term is highlighted with a blue background. On the right, a text snippet is shown with two phrases highlighted in yellow: 'networks are interactive' and 'interactive social networks'. The text snippet reads: 'Does China still have the means to completely control chinese Internet users ? Information access is now free in China. With the rapid emergence of micro-blogging services, authorities seem to be confronted with growing difficulties. And the fact that these social networks are interactive prevents China from controlling public opinion. That's the analysis of Hu Yong, a journalism professor at Beijing University. When interviewed by AFP he drew a parallel with the insurrections that struck several Arab countries. As in Tunisia, Egypt or Libya, it is through these interactive social networks that popular movements are formed. Such websites have become a key feeding ground for public opinion. The control of the Internet and virtual neutering of all online information sources, including through aggressive means, are the only way China has succeeded in maintaining the stability of its political regime. The emergence of interactive social networks has the potential of disturbing this stability by liberating the voice of the Chinese people... and feeding social protest movements.'

# Relevance Scoring : before & after

Before  
18 Concepts

Extraction from: STFKnowledge

Does **China** still have the means to completely control chinese **Internet** users ? **Information access** is now free in **China**. With the rapid emergence of micro-blogging **services**, authorities seem to be confronted with growing difficulties. And the fact that these social **networks are interactive** prevents **China** from controlling **public opinion**. That's the analysis of Hu Yong, a **journalism professor** at Beijing **University**. When interviewed by AFP he drew a parallel with the **insurrections** that struck several Arab **countries**. As in **Tunisia**, **Egypt** or **Libya**, it is through these **interactive social networks** that popular movements are **formed**. Such **websites** have become a key feeding ground for **public opinion**. The control of the **Internet** and virtual neutering of all online **information sources**, including through aggressive means, are the only way **China** has succeeded in maintaining the stability of its **political regime**. The emergence of **interactive social networks** has the potential of disturbing this stability by liberating the voice of the Chinese people... and feeding **social protest movements**.

After  
8 concepts

Extraction from: STFKnowledge

Does **China** still have the means to completely control chinese **Internet** users ? **Information access** is now free in **China**. With the rapid emergence of micro-blogging services, authorities seem to be confronted with growing difficulties. And the fact that these social **networks are interactive** prevents **China** from controlling **public opinion**. That's the analysis of Hu Yong, a journalism professor at Beijing University. When interviewed by AFP he drew a parallel with the **insurrections** that struck several Arab countries. As in **Tunisia**, **Egypt** or **Libya**, it is through these **interactive social networks** that popular movements are formed. Such websites have become a key feeding ground for **public opinion**. The control of the **Internet** and virtual neutering of all online **information sources**, including through aggressive means, are the only way **China** has succeeded in maintaining the stability of its political regime. The emergence of **interactive social networks** has the potential of disturbing this stability by liberating the voice of the Chinese people... and feeding **social protest movements**.

# Multilingual taxonomy benefits

## EuroVoc

- ▼ (-)
  - ENERGY
    - ▼ (-)
    - soft energy
      - renewable energy (35)
      - solar energy (12)
      - biogas (6)
      - wind energy (6)
    - ▷ energy policy (11)
    - nuclear industry (4)
  - ▷ FINANCE (14)
  - ▷ GEOGRAPHY (13)
  - ▷ BUSINESS AND COMPETITION (12)
  - ▷ ENVIRONMENT (11)
  - ▷ INDUSTRY (10)
  - ▷ TRANSPORT (9)
  - ▷ ECONOMICS (8)
  - TRADE (8)
  - ▷ POLITICS (7)
  - ▷ PRODUCTION, TECHNOLOGY AND RESEARCH (7)
  - SOCIAL QUESTIONS (6)
  - EMPLOYMENT AND WORKING CONDITIONS (4)
  - EUROPEAN COMMUNITIES (4)
  - LAW (3)
  - ▷ SCIENCE (3)
- Show fewer

Home

## Trains run like, and on, the wind

Tue, 09/13/2011 - 10:00 — admin

Trains run like, and on, the wind

It will not be easy to run a national railroad on **renewable energy** like wind, hydro and solar power,...

**ERIK KIRSCHBAUM** 09/13/2011 It will not be easy to run a national railroad on **renewable energy** like wind, hydro and solar power, but that is what Deutsche Bahn of **Germany** aims to do, for one simple reason: It is what consumers want. Deutsche Bahn says it wants to raise the percentage of wind, hydro and **solar energy** used in powering its trains from 20 percent now to 28 percent in **2014** and to become carbon-free **by 2050** — targets that exceed the **German government's** already ambitious national goals. "Consumers in **Germany** have made it clear they want us all to get away from nuclear **energy and to more renewable** energy," said **Hans-Jürgen Witschke**, chief executive of Deutsche Bahn Energie, which supplies electricity for trains in **Germany**. "It's what customers want, and we're making it happen," **Mr. Witschke** said during an interview. "The demand for green **electricity** keeps rising each year, and that'll continue." Prevailing attitudes in **Germany** were already decidedly green before the accident at the **Fukushima** Daiichi nuclear complex in **Japan**, set off by the earthquake and tsunami on **March 11**. After the nuclear crisis in **Japan**, the **Berlin government** abruptly reversed course on nuclear

## Related documents

- Wir wollen den Atomausstieg
- Les énergies renouvelables a la conquete des reseaux mobiles
- China set to challenge global wind industry
- Waste problem for green cars: What to do with the batteries
- Die Rechnung kommt noch

Hide highlight

# Conclusion

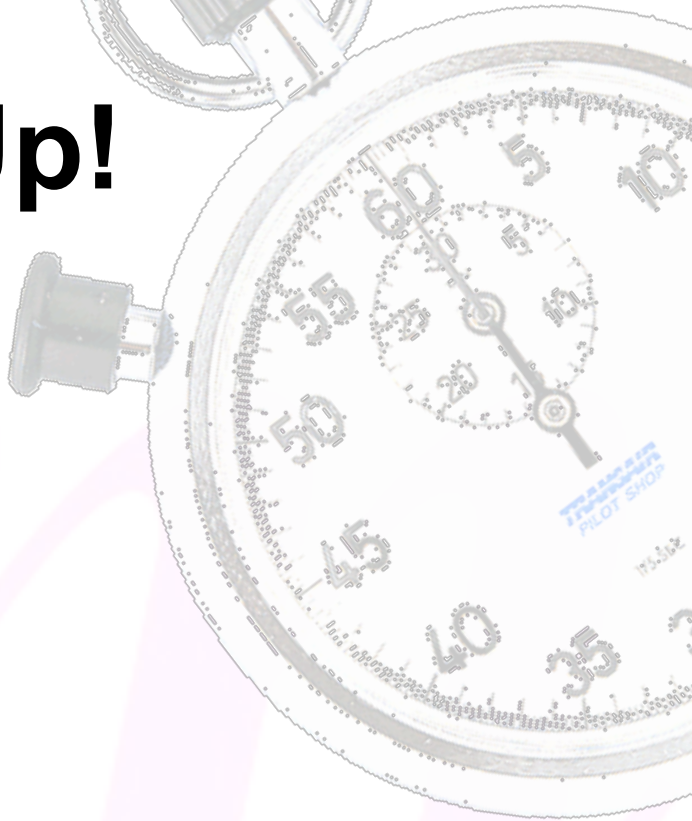
## A **platform** for present and future growth

- Compelling content – serves readers and authors
- Innovative products & formats
- Editorial Productivity

## **Get started now !** With a pragmatic approach

- Easy step into semantic content enrichment
- Don't start from scratch
- Fast time-to-value

# Time's Up!



## About your speaker:

**Name:** Daniel Mayer

**Company:** 

**Tel:** +33 1 8098 1145

**Email:** [daniel.mayer@temis.com](mailto:daniel.mayer@temis.com)

**Social Media:** @TEMIS\_Group