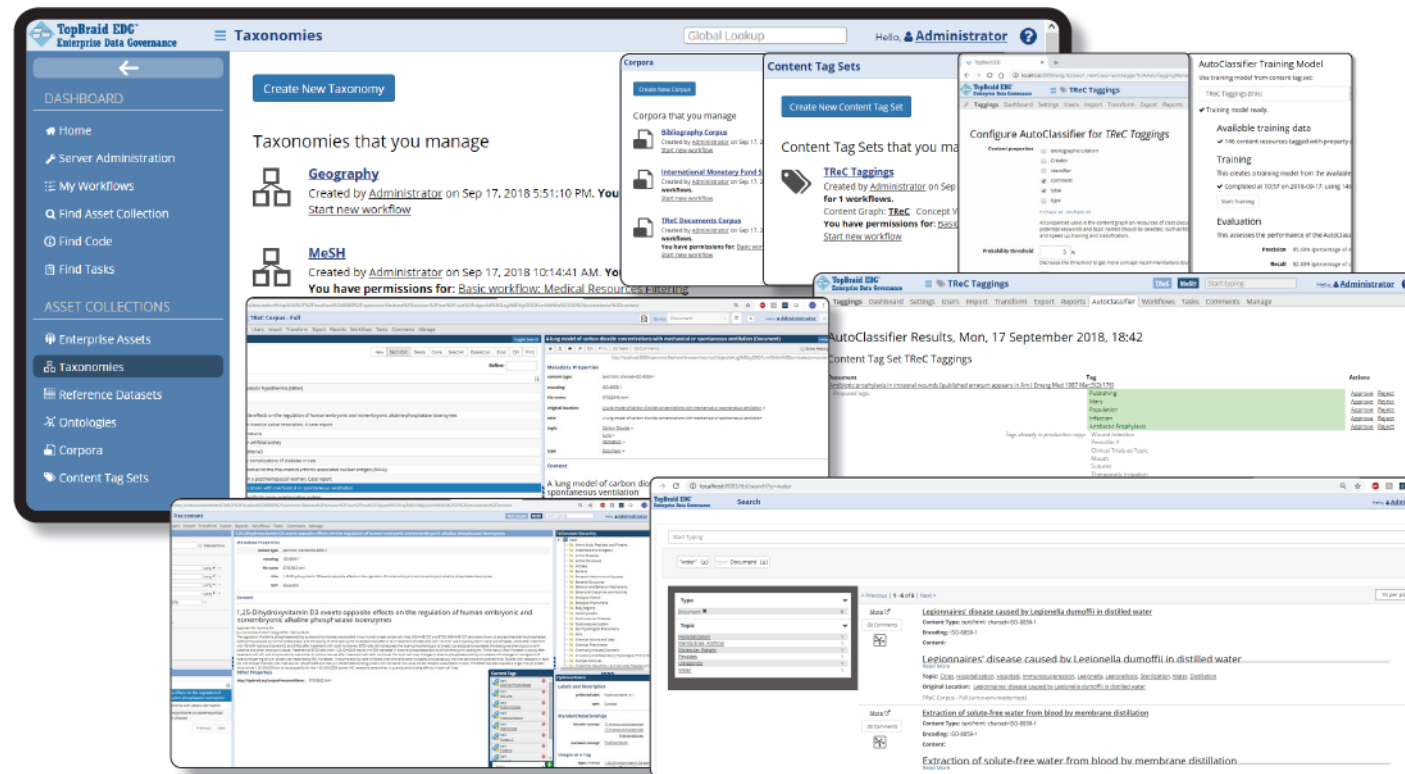


Guided ML: Intelligent Autoclassification of Content Using Managed Vocabularies

TopBraid Tagger's auto-classification capability delivers tagging results that are as consistent and accurate as those created by content indexing experts using typical domain taxonomies.



June 27, 2019 11:30am EDT

Today's Program

- **INTRODUCTION**
 - Discuss unstructured content in the enterprise, tagging and semantic search
- **DEMO**
 - Demo of these capabilities using Tagger and AutoClassifier as part of TopBraid EDG
- **WRAP-UP**
- **Q & A**



Jesse Lambert



Nuno Lopes

FOUNDATION

- TopQuadrant was founded in 2001
- Strong commitment to standards-based approaches to data semantics

MISSION

- Empower people and drive results — by making enterprise information meaningful



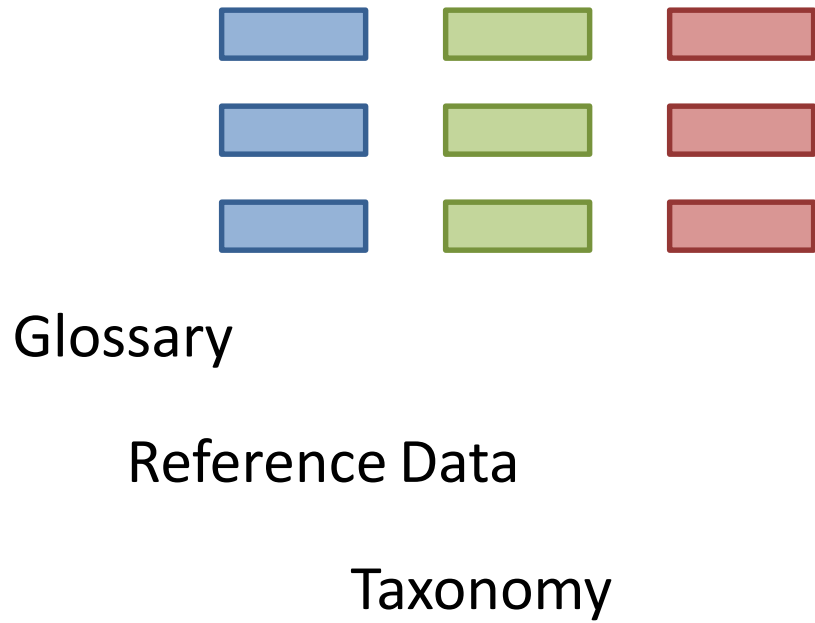
FOCUS

- Provide comprehensive data governance solutions

Unstructured Content in the Enterprise...

- Keyword/Phrase/Conversation search doesn't work so well.
 - Domain specific and technical language.
 - Rules are complex and time consuming to develop.
 - Machine learning approaches require large training sets.
- Different kinds of objects (slides, emails, documents...etc).
 - Objects are all over the place.
- **Up to 80% of enterprise information is unstructured.**

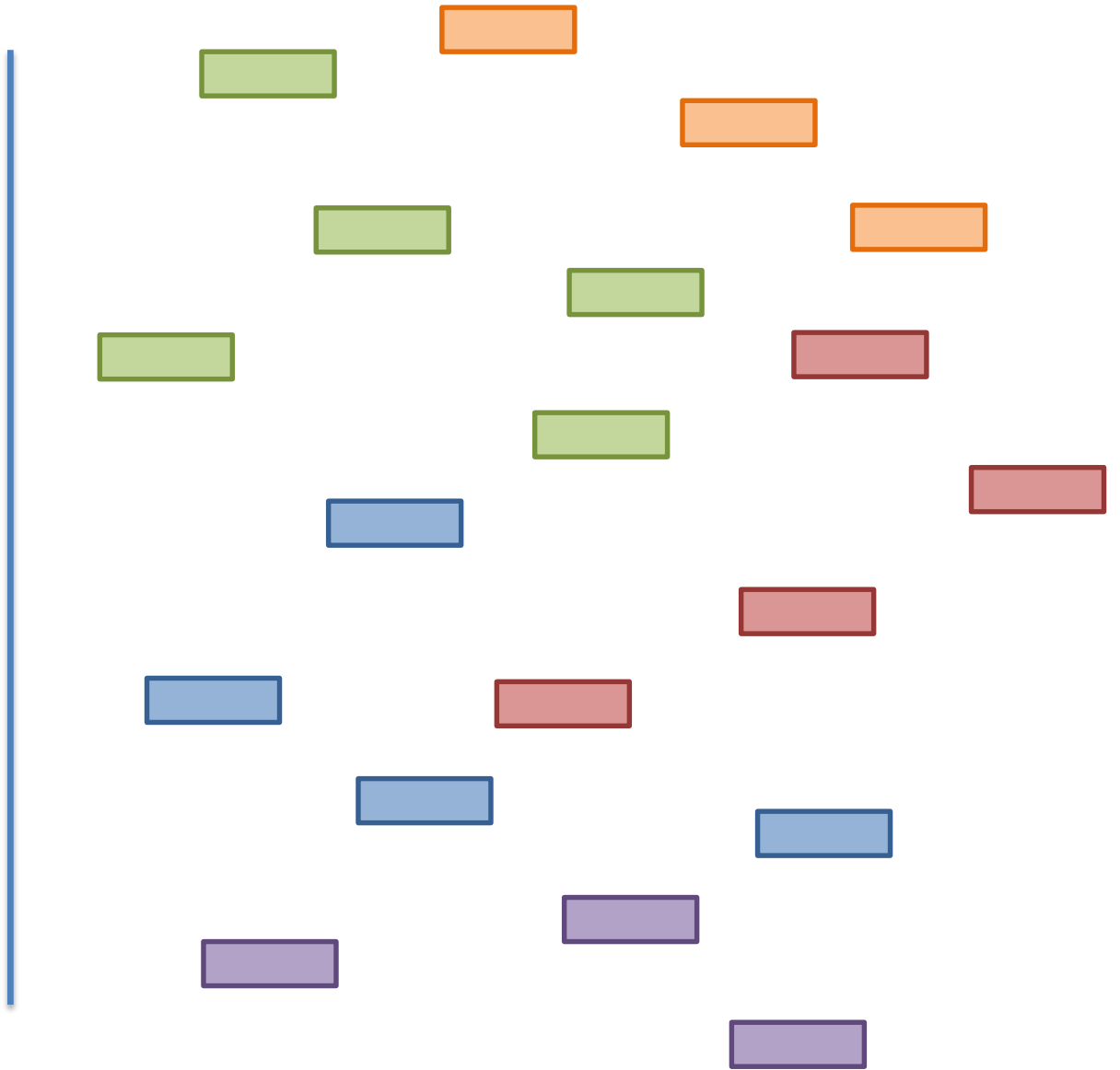
Bridging the Gap



20 % Structured

vs

80% Unstructured

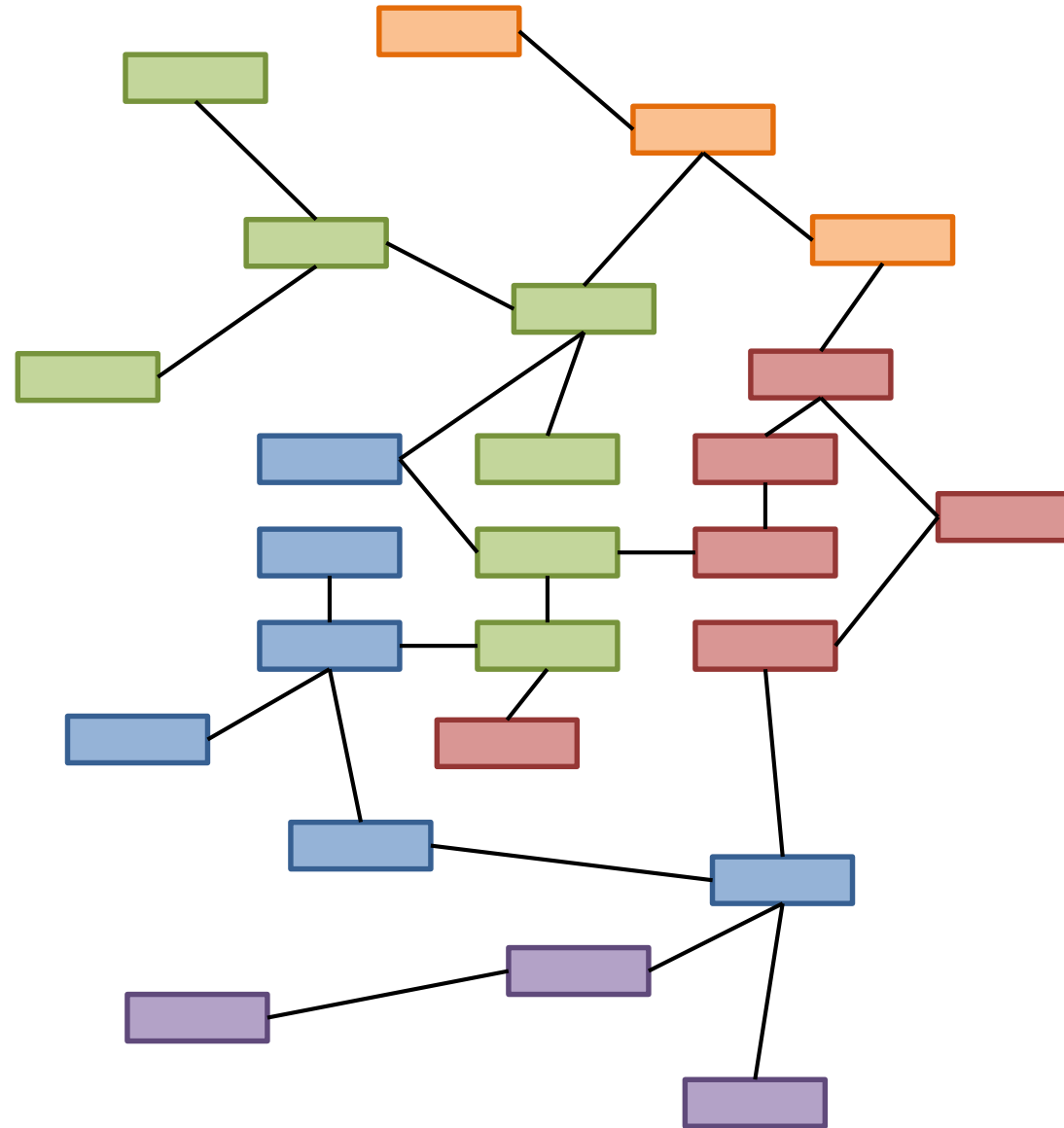


Bridging the Gap

Add value to more content
faster and at a larger scale

Machine learning—no need
to manually create rules

Human in the loop for review



Governed process

Integrate with other
systems via APIs

TopBraid EDG: Overview

TopBraid EDG offers prebuilt models and features that support integrated governance over a complete range of information assets and governance needs, using enterprise knowledge graphs to capture semantics

TOPBRAID EDG - GOVERNANCE PACKAGES



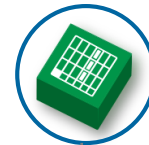
Vocabulary Management

lets organizations create, connect and use taxonomies and ontologies



Metadata Management

lets organizations govern technical metadata



Reference Data Management

makes it easy to bring consistency and accuracy to the use of reference data across enterprise applications



Business Glossaries

provides a simple starting place for a data governance initiative

TOPBRAID EDG – ADD ON MODULES

TopBraid Explorer lets a broader community of data stakeholders explore information governed by TopBraid EDG

Customer Defined Asset Collection Type is an add-on module that can be added to any of the packages within TopBraid EDG.

TopBraid Tagger and AutoClassifier links controlled vocabulary terms to content. Content resources can be tagged manually or auto-tagged using Tagger's AutoClassifier capabilities.

TopBraid Data Platform provides high availability TopBraid servers for continuous operation of business functions by replicating data across a cluster of servers

TopBraid EDG: Vocabulary Management Package

Initially offered as a separate product called TopBraid Enterprise Vocabulary Net(EVN) until 2017, now offered as one of the packages within EDG: TopBraid EDG-VM (Vocabulary Management)

Taxonomies

- Based on SKOS or SKOS-XL

Ontologies

- Expressed in SHACL and/or OWL

Crosswalks

- Contain many-to-many links between two asset collections with similar terms
- Created manually or automatically

Corpora

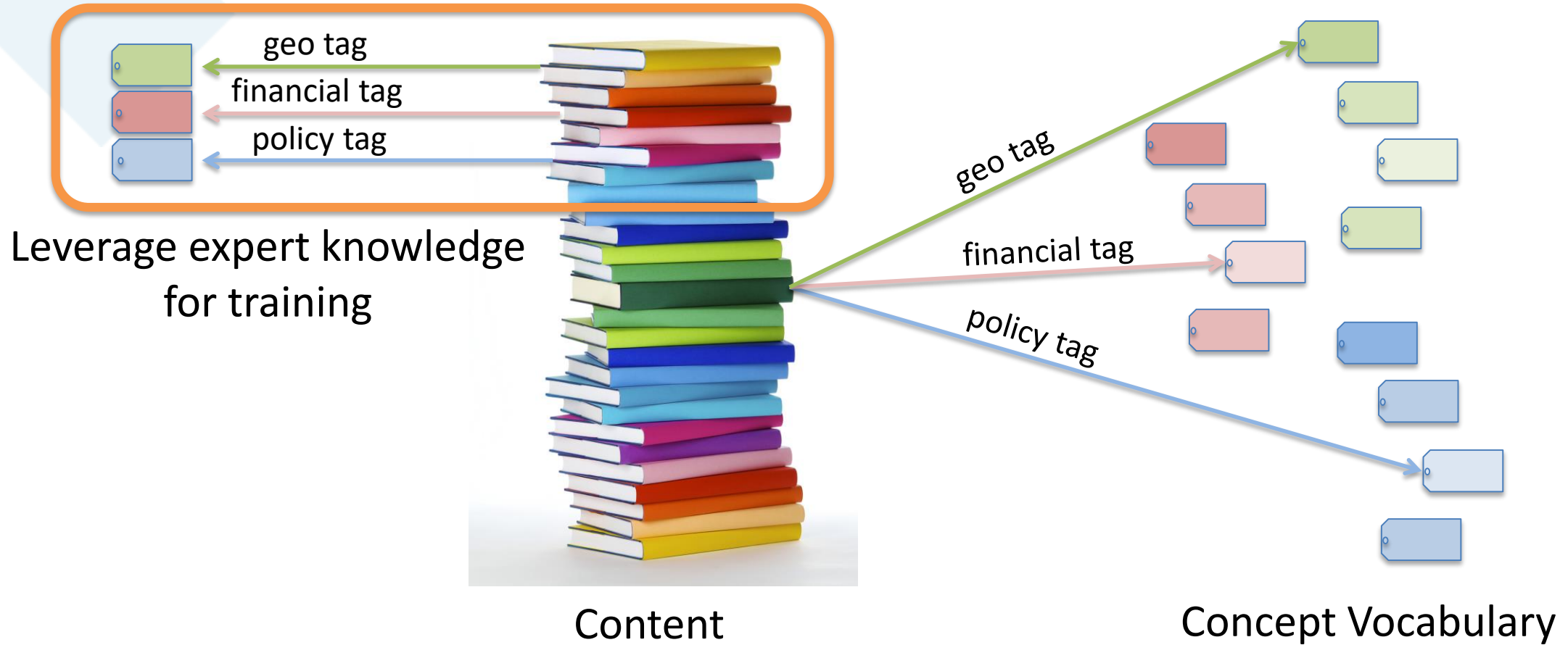
- Controlled editing, provide access to documents in external document/content management systems
- All common document forms are supported e.g., MS Office, PDF, etc.

Content Tagsets

- Connect (or 'tag') documents with terms from controlled vocabularies
- Created manually or automatically

This package is often used together with the Tagger and AutoClassifier add-on module which adds two more asset collections:

Tagging



Semantic Search over published Macroeconomics Research Reports



Search...

« Previous | 1 - 10 of 5216 | Next »

10 per page ▾ By label ▾

Limit your search

report type	
ECB European Central Bank Working Paper	1775
FRB International Finance Discussion Papers	1155
Banque De France Working Paper	555
Banco de España Working Paper	513
Bank for International Settlements Working Papers	492
FRB Finance and Economics Discussion Series	428
BoE Staff Working Paper	308

year >

quantity kind term >

process term >

amendment term >

strategy term >

Add property:

Accounting for Spanish productivity growth using sectoral data: New Evidence

Banco de España Working Paper 0110

In this paper we measure, at sectoral level, the sources of Spanish productivity growth, distinguishing among the roles played by labor productivity, the degree of factor substitution and total factor productivity (Solow residual). In terms of value added, total factor productivity growth in the manufacturing sector is higher than in services (2.4% and 1.1% respectively), although the Communication Sector constitutes an exception. In terms of gross production productivity growth in both sectors is virtually the same (0.6% and 0.7% respectively) because of the relative lower weight of intermediate inputs in producing services. During the latest cyclical expansion, 1995-1999, we have observed a marked deceleration in labor productivity, associated to an increase in employment, and to a deceleration in the capital-labor ratio, while total factor productivity has remained more stable.

url reference: <http://www.bde.es/f/webbde/SES/Secciones/Publicaciones/PublicacionesSeriasdas/DocumentosTrabajo/01/Fic/dt0110e.pdf>

Rational Underdevelopment

Banco de España Working Paper 0114

We propose a two-region two-sector model of uneven development, where technological change benefits either the lagging or the leading region. In this framework inter-regional transfers may lead to persistent underdevelopment, by raising wages without changing productivity, transfers reduce the chance of the backward region adopting a new technology and taking off. Due to uncertainty about which region benefits from technological change, the backward region may rationally choose to remain underdeveloped, while the advanced region continues to pay transfers. The Model provides a rationale for cases, such as Italy's Mezzogiorno, where the same rich region subsidizes the same poor region on a continuous basis.

url reference: <http://www.bde.es/f/webbde/SES/Secciones/Publicaciones/PublicacionesSeriasdas/DocumentosTrabajo/01/Fic/dt0114e.pdf>

The Transmission Mechanism of European Monetary Policy: Is There Heterogeneity? Is It Changing Over Time?

Banco de España Working Paper 0115

The European Central Bank (ECB) has already moved interest rates several times since it started to operate in January 1999 and yet we don't know what the magnitude and timing of the effects of its actions actually are. What are the effects on prices and output of a change in the common short-term interest rate? How long do these effects take to materialize? Are there differences in the impact across European countries and regions? Are these differences changing over time? We propose to overcome some of these difficulties by rephrasing some of the questions above in the framework of a dynamic heterogeneous panel data model recently proposed by Canova and Ciccarelli (2000).

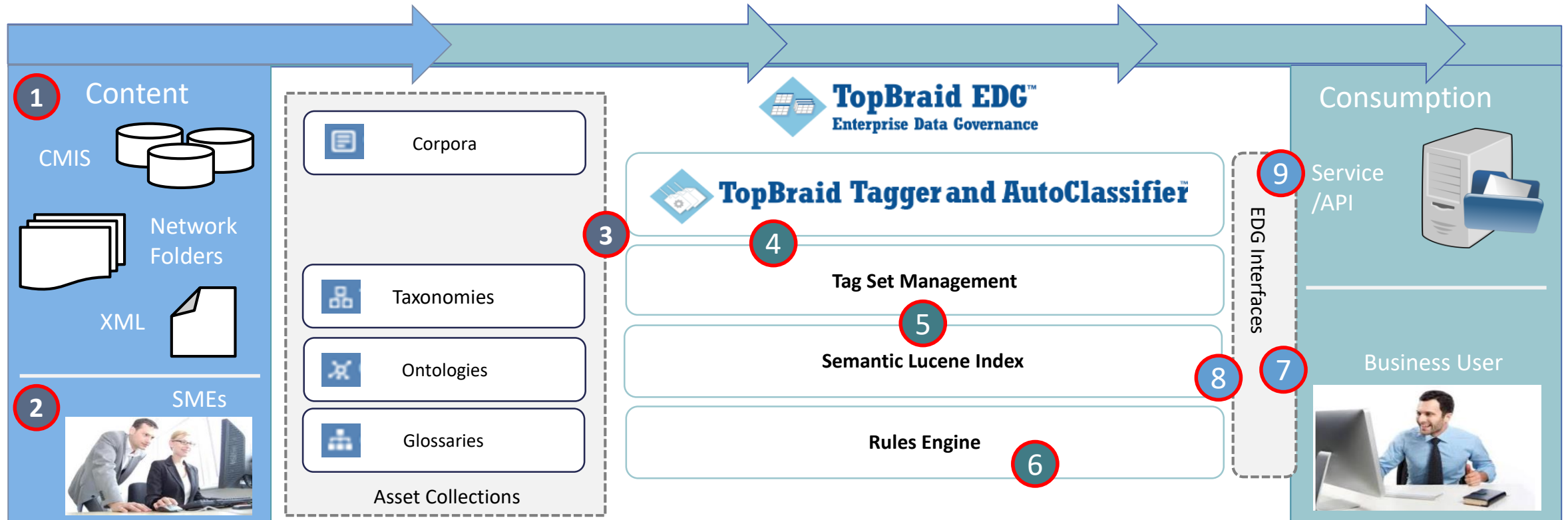
url reference: <http://www.bde.es/f/webbde/SES/Secciones/Publicaciones/PublicacionesSeriasdas/DocumentosTrabajo/01/Fic/dt0115e.pdf>

MAECO Taxonomy

- This taxonomy of over 7000 terms was built from several authoritative sources of terminology. Key contributions to the taxonomy were:
 - **U.S. Commodity Futures Trading Commission (CFTC) glossary,**
<http://www.cftc.gov/ConsumerProtection/EducationCenter/CFTCGlossary/index.htm>
 - **Professor Deardorff's glossary, University of Michigan,**
<http://www-personal.umich.edu/~alandear/>
 - **The Bank of England's (BoE) Prudential Regulation Authority (PRA) glossaries,**
<http://www.bankofengland.co.uk/pru/Pages/policy/handbook.aspx>
 - **The Journal of Economic Literature (JEL) taxonomy for Linked Open Data,**
<http://zbw.eu/labs/en/project/jel-lod>
 - **A financial taxonomy from TopQuadrant**
 - **The Financial Industry Business Ontology (FIBO),**
<https://www.fibo.com/en/>

DEMO

TopBraid EDG: Search Enrichment



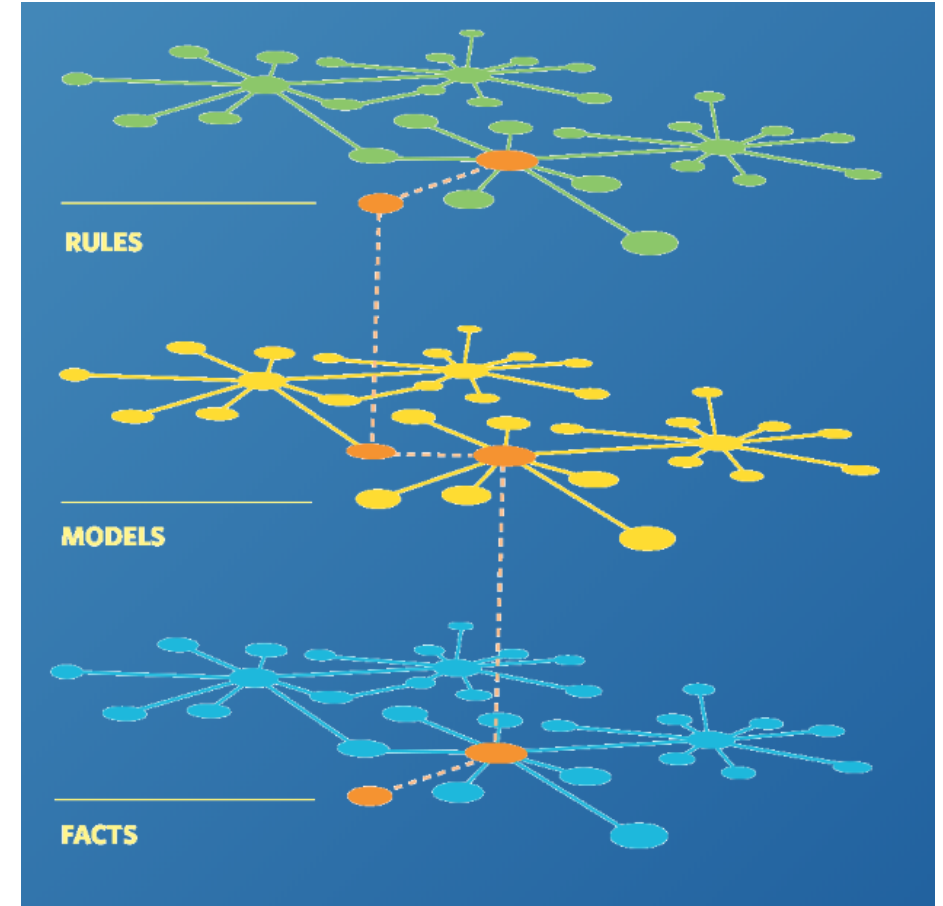
- Content is retrieved/received over many connectors. SMEs of all sorts develop/integrate vocabularies. As needed, they also review and curate auto-classification results providing input that improves results accuracy over time.
- Vocabularies and curation feedback are leveraged by AutoClassifier to extract information from unstructured content inside Asset Collections.

- AutoClassifier uses machine learning to populate maintainable Tag Sets (relationships between 'Content' and 'Concepts').
- Text, Tags, and other relevant characteristics of Assets are indexed for optimized search and discovery.
- Semantic models and business rules can be used to further post process and extend results

- Search user submits a request and the interface sends the request on to the query engine.
- The query is interpreted and semantic search is executed. Results include relevant metadata for customized user experience.
- Similar to the search user, applications submit requests to an EDG interface. Determined by the interface (Web Service, Saved Search, SPARQL endpoint), results are prepared and sent in appropriate format.

Benefits of a Knowledge Graph based Platform for Data Governance 2.0

- Flexibility and extensibility based on standards
- Integration of reasoning and machine learning
- Enabling people (UI) and software (APIs/web services) to view, follow and query
- Bridging of data and metadata “silos” to provide seamless data governance
- Delivery of Knowledge-driven data governance



As an enterprise knowledge graph infrastructure, TopBraid EDG supports Data Governance 2.0 and applications of AI / ML



... Questions?

To Learn More about TopBraid Tagger and AutoClassifier

EDG Product Info:

- <https://www.topquadrant.com/products/topbraid-enterprise-data-governance/>
- <https://www.topquadrant.com/products/topbraid-tagger-autoclassifier/>

Contact us: at info@topquadrant.com to:

- Discuss data governance solutions
- Request a more targeted demo of Tagger and AutoClassifier
- Request a more targeted demo of TopBraid EDG
- Ask for a free EDG evaluation account

More Resources ...

More Webinar Recordings, Slides, Q&A:

- <https://www.topquadrant.com/knowledge-assets/topquadrant-webinars/>

Short Videos:

- TopBraid EDG “Quick Grok” Videos
<https://www.topquadrant.com/knowledge-assets/videos/>
- TopBraid EDG Animated Video
<http://www.topquadrant.com/EDGvideoemail/>

Blog:

- <https://www.topquadrant.com/the-semantic-ecosystems-journal/>

Data Governance White Papers

- <https://www.topquadrant.com/knowledge-assets/whitepapers/https://www.topquadrant.com/knowledge-assets/whitepapers/>