



**LT-Innovate
Solution Proposal
For BMS**

Leverkusen, 5 December 2013

Anna Sidorova, Vera Kuzmina, Head of Marketing, Head of Innovations & Business Development Manager, ABBYY

Luc Meertens & Joachim Vandenbogaert
CEO & Machine Translation Specialist, CROSSLANG

Kurt Eberle, Managing Director, Lingenio

Pierre Bernassau, Director, Client Services EMEA, Systran

Christina Niegel & Manfred Pitz, Pre-sales Consultant & Director Sales Germany, TEMIS

Jasper Dencker-Hahn, CEO, VETAIL-X

Philippe Wacker & Andrew Joscelyne, Secretary-General & Senior Advisor, LT-Innovate



LT-Innovate

A top-down photograph showing approximately 15 hands of various skin tones (ranging from light to dark brown) pressed together in a circle on a sandy surface. The hands are arranged in a ring, with fingers pointing towards the center. The sand is light-colored and textured. The overall composition is circular and centered.

**180 LT vendors
in 25 countries**

Yearly LT-Innovate Summit (June) & LT-Innovate Award



To explore strange new worlds, to seek out new data and new knowledge, to boldly go where LT has not gone before...



<http://bit.ly/13UAaST>

Discover the world of
Language Technologies:
Download LT2013!

Workshops



INNOVATION BROKERAGE

Improve framework conditions

Facilitate
technology
transfer

Help vendors
collaborate

Help buyers
define their needs



Researchers

Vendors

Buyers

Procure
applied research

Help vendors
increase their market

Innovation Brokerage



Bayer MaterialScience

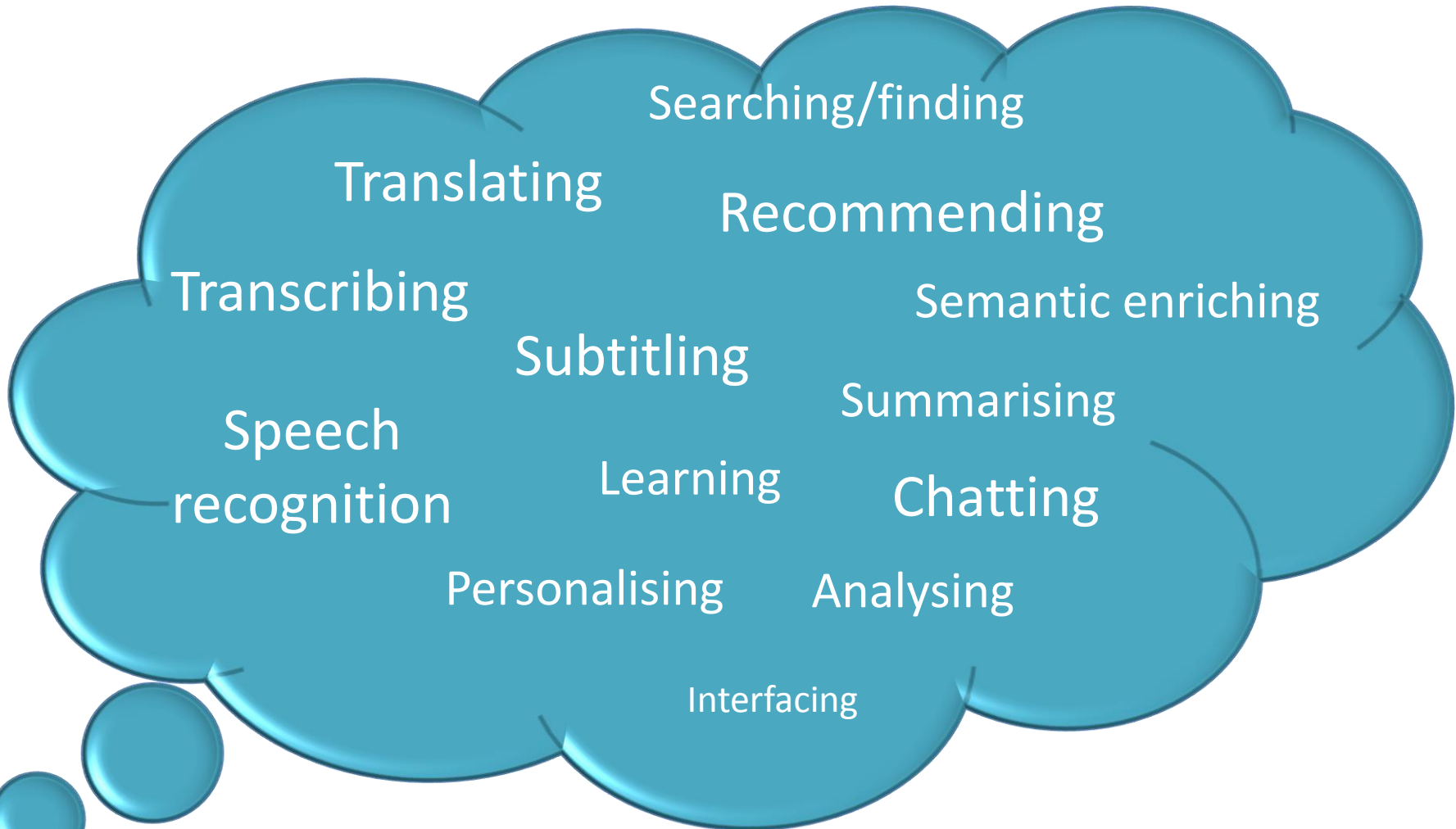
Language Technology

is the Drill
that extracts

the meaning
of Big Data!



A SECURE, EXTENDABLE CLOUD CONTAINING TOOLS & APPLICATIONS FOR:



To build a solution, we need to consider:

- **Leveraging BMS legacy translation/language data**
- **Integrating seamlessly with the BMS CMS and IT and Unified Comms platform**
- **Working with BMS *people* as intelligently as possible**

It's 2015.

Marlene Meyer works
at BMS as a PUR
researcher.

She has a busy day
ahead.

Here is her to-do list ...



MARLENE MEYER'S DIARY

To-Do list for week 05

Watch CEO's message
(Must understand it in detail to inspire team!)

Talk with Fen Shiu about his team's work

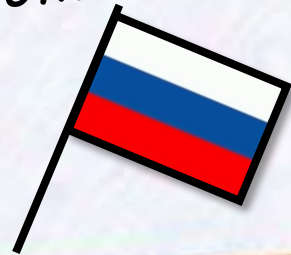


(Chinese? Ouch!)

NB: Need to find the latest research on current project
Lots of documents, only a few are relevant...

Read and summarise recent reports from Russian team!

URGENT!!



ACT 1 – A WORD FROM THE CEO

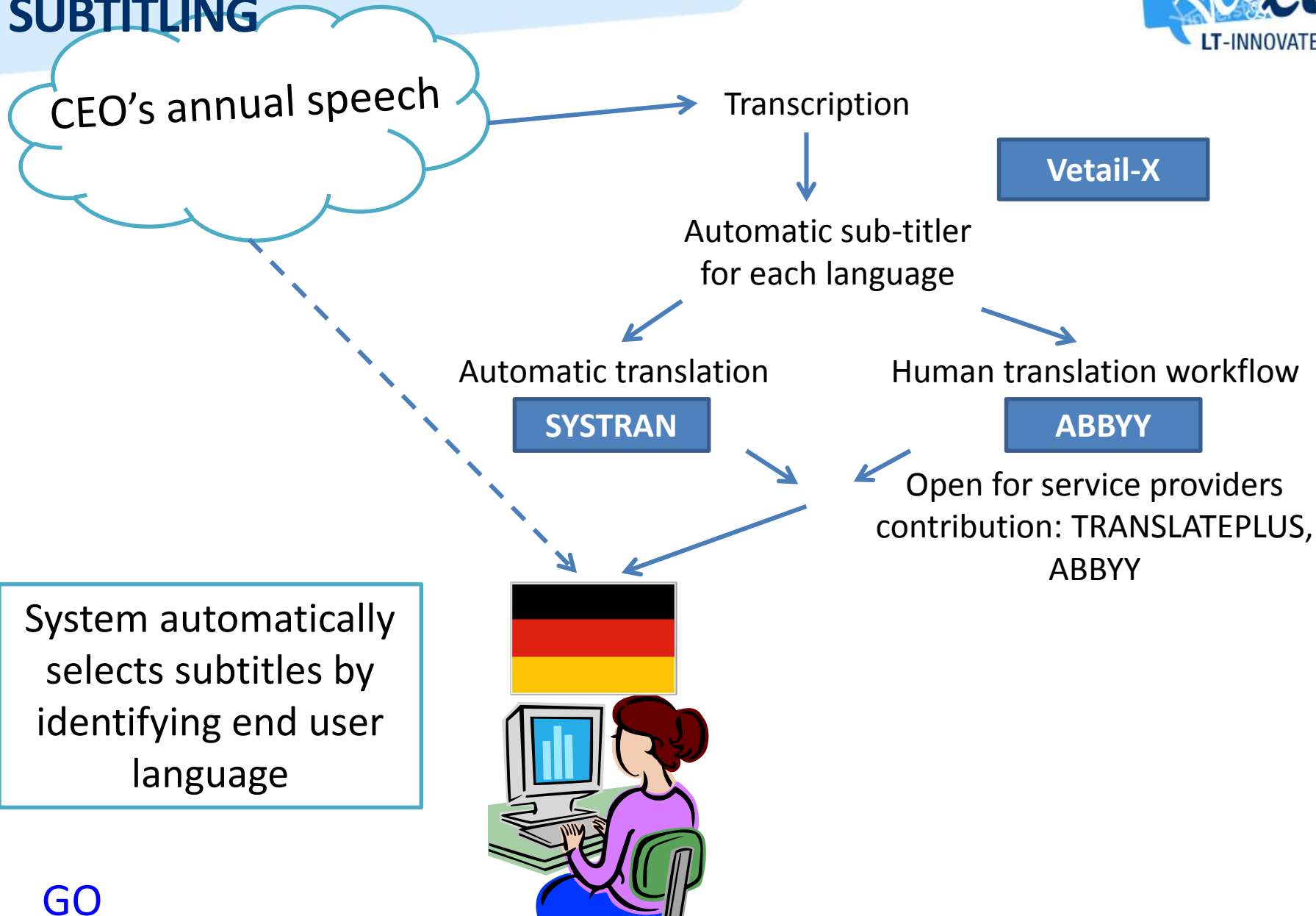
ACT 2 – YOU ARE WHAT YOU SEARCH FOR

ACT 3 – GLOBAL CHATTERBOX

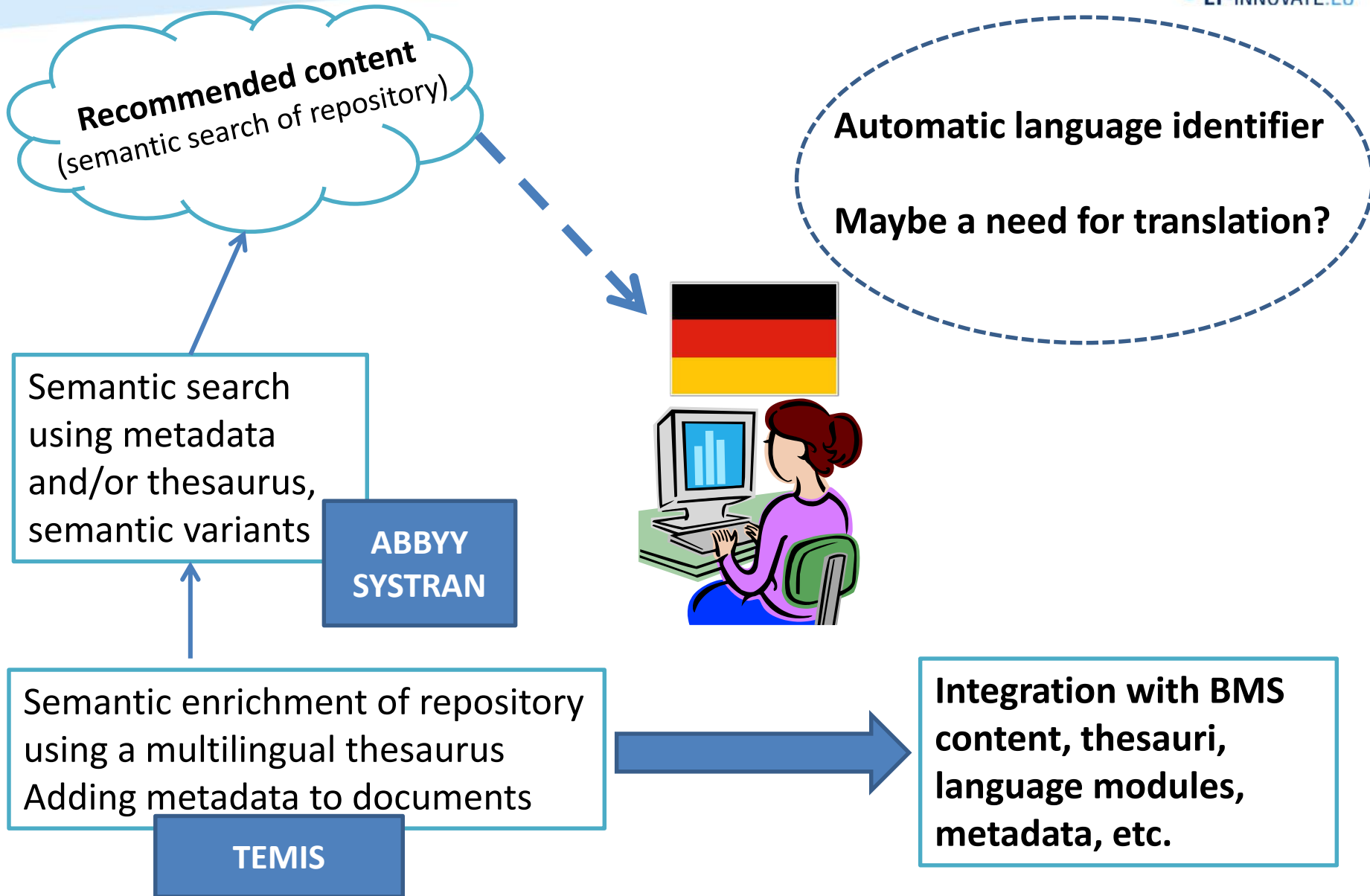
ACT 4 – FOUND IN TRANSLATION

FINALE – THE WHITE BOX

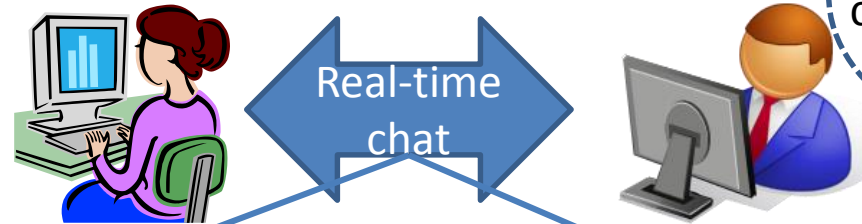
ACT 1 – MULTILINGUAL SUBTITLING



ACT 2 – YOU ARE WHAT YOU SEARCH FOR



Integrate with BMS unified communications system



Real time chat and consecutive Q&A support

Written chat

Spoken chat

Automated workflow

Human translation workflow with SLA

Automated workflow
Speech recognition + MT

Human Interpretation

SYSTRAN

ABBY

**SEDAT 85
SYSTRAN
LINGENIO**

ABBY

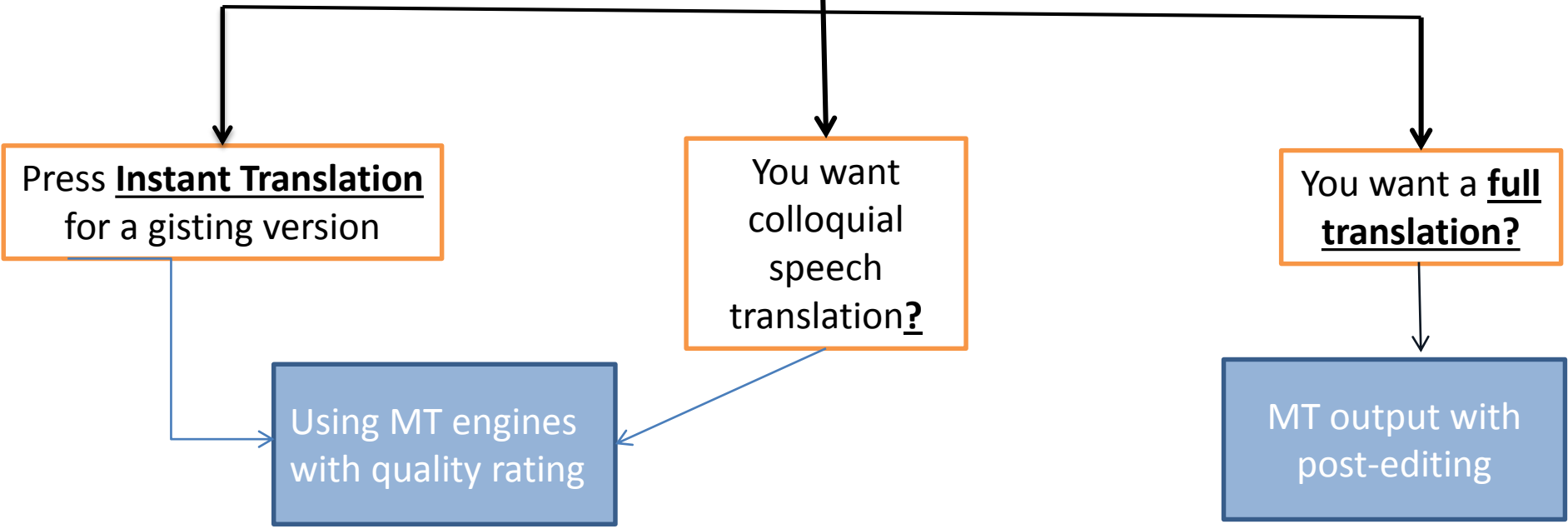
Issues

- Colloquial language
- Technical domain language
- Extending service to other needs, e.g. meetings

ACT 4 – FOUND THROUGH TRANSLATION



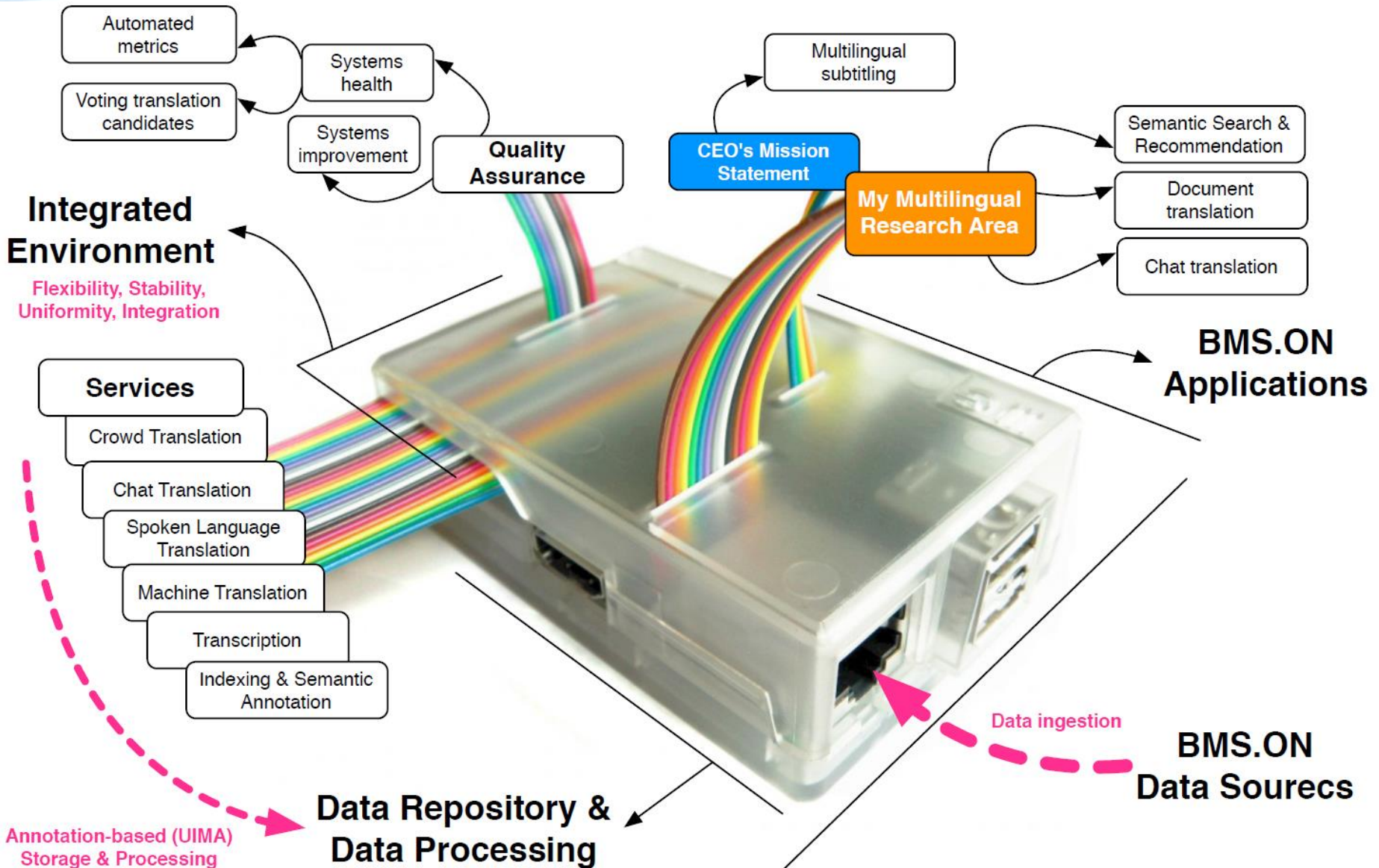
Automatic language identifier:
Do you need a translation?



ABBY
LINGENIO
SYSTRAN

TRANSLATE PLUS

FINALE – THE WHITE BOX



Thank you!

- What types of linguistic assets does BMS have, where are they stored and how are they being reused? (TMs, glossaries, dictionaries)
- Details on the existing process of localization from ordering to reception
- Share of 11 languages in total content?
- How many users would you expect for a translation service?
- Other types of document translation
- Email etc. translation

OUR QUESTIONS ABOUT CONTENT SEARCH & THE REPOSITORY



- Who has implemented the BMS.ON platform? Bayer Business Services or an external system integrator?
- Is data both internal and from external sources?
- Which version of Microsoft SharePoint is currently being used for BMS.ON?
- What kind of (SharePoint) search engine and which version of it is being used? Is it Microsoft FAST?
- Is there already any controlled vocabulary (thesaurus, ontology, list of words) that can be utilized?
- Documents:
 - ✓ overall volume of the database?
 - ✓ how many new documents added on average per day?
 - ✓ data formats?

OUR QUESTIONS ABOUT CONTENT SEARCH & THE REPOSITORY



- What is the share of each 11 languages to the whole amount of data?
- What are the sources where data comes from (if not stored all the time in the repository)?
- Data comes only from internal BMS sources only? Or external data will be used, too?

OUR QUESTIONS ABOUT MULTILINGUAL CHAT



- What platform does BMS currently use for communication, chatting (Lync?)?
- Do you target chat rooms with n2n discussions.
- What types of communication does BMS have/consider useful? (forums, knowledge bases, corporate social networks, etc.). Is it communication one to one, one to many, many to many?
- Do you use phone conferencing or a simple one to one discussion?
- Does BMS consider only real-time communication or consecutive communication as well? (if so, does BMS have any multilingual Q&A queues?)

OUR QUESTIONS ABOUT VIDEO SUBTITLING

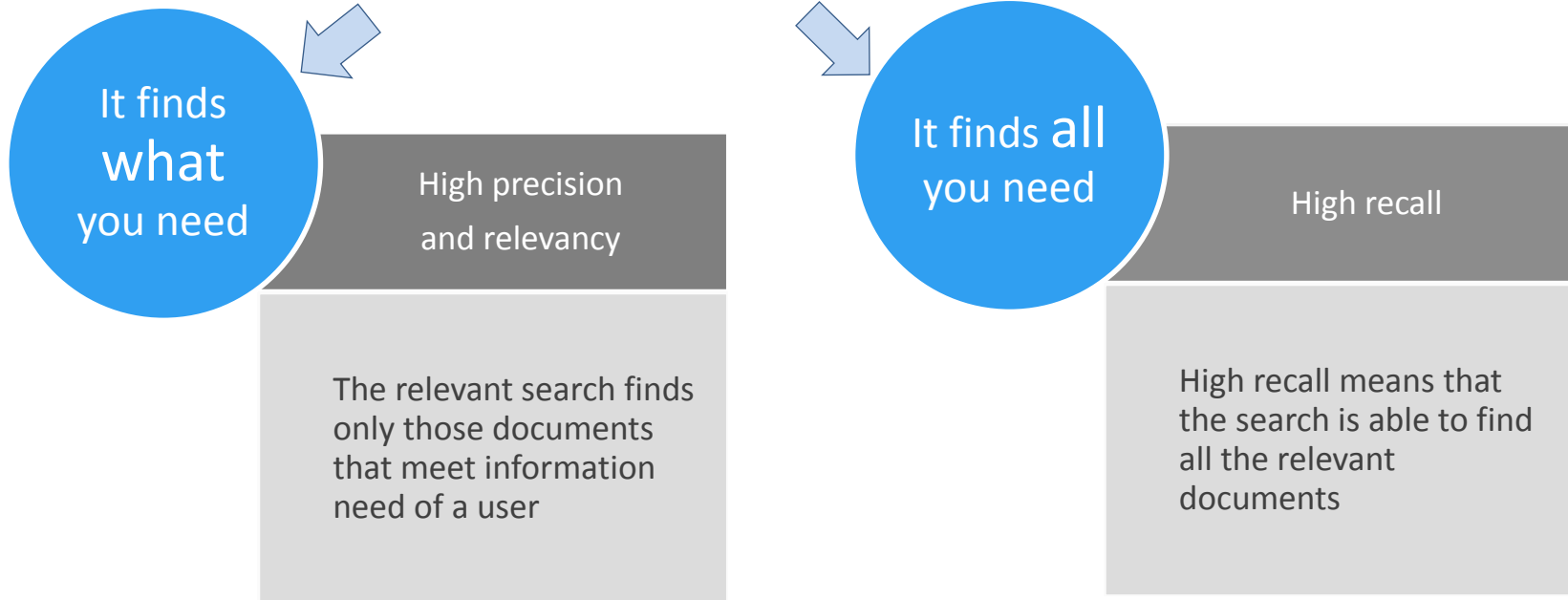


- How are videos delivered to the user (Web Interface, Stand-alone player, etc.)?
- What is the estimated amount of video data?
- What languages are the videos in?
- What kinds of videos are to be subtitled? (CEO speech, product videos, tutorials, UGC)



SEARCHING AND FINDING NEEDED DOCUMENTS

Search is successful when it finds what
the user is searching for.
There are two criteria:



SEARCH SCENARIO EXAMPLES

SCENARIO 1: SEARCH FOR 1 WORD –
SINGLE MEANING OR MULTI-MEANING.

German researcher looks for the word
POLYURETHAN

Semantic metadata enrichment for multilingual thesauri will help getting high relevance and **finding multilingual documents containing needed terminology in 11 languages**

SCENARIO 2: SEARCH FOR PHRASE.

German researcher looks for **GEBRAUCH VON POLYURETHAN IN GOLFBÄLLEN**

- Verwendung von Polyurethan in Golfbällen
- Studie über Polyurethanegebrauch in Golfbällen
- die Golfbälle werden von Polyurethan hergestellt
- Polyurethan in Bällen für Golf

All phrases with the **same meaning will be found**

SCENARIO 3: SEARCH FOR ACTIONS IN THE CONTEXT.

German researcher looks for **STOFFE VERWENDET FÜR GOLFBÄLLEN-HERSTELLUNG**

All documents containing **any words with the semantic identification MATERIAL** in the connection of production golf balls will be found



User

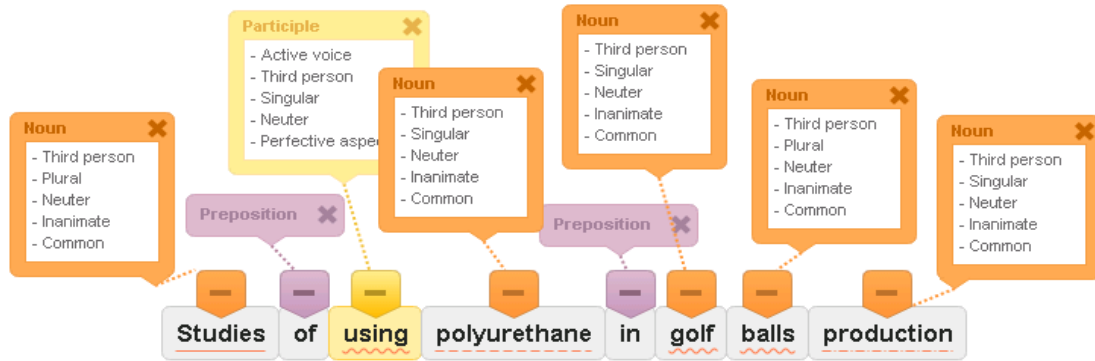
SEMANTIC SEARCH APPROACH: HOW IT WORKS

ACCURATE TEXT ANALYSIS AND HIGH RECALL

Through exhaustive semantic and syntactic parsing, Semantic Search identifies the meaning of the words in a text and relations among them.

- | | |
|--|--|
| 1. Use of Polyurethan in golf balls | Verwendung von Polyurethan in Golfbällen |
| 2. Research on using Polyurethane in golf balls | Studie über Polyurethangebrauch in Golfbällen |
| 3. making golf balls from polyurethane | die Golfbälle werden von Polyurethan hergestellt |
| 4. polyurethane in balls for golf | Polyurethan in Bällen für Golf |
| 5. other material used for golf balls production | Andere Stoffe gebraucht in Golfbällenherstellung |

The same meaning



study: TO_LEARN_AND_RESEARCH
COGNITION_ACTIVITY
VERBS_OF_ACTIVITY
SITUATION

use: use
TO_USE
VERBS_OF_ACTIVITY
SITUATION

polyurethane: POLYURETHANE
POLYMER
ARTIFICIAL_SUBSTANCE
SUBSTANCE

ball: BALL
SPORT_DEVICE_FOR_THROWING,
SPORT_DEVICE
DEVICE

production: production
TO_MANUFACTURE
TO_RELEASE_AS_TO_CREATE
TO_MAKE

Semantic classes

Multilingual Thesaurus

- build / extend / update a **domain-specific thesaurus** covering relevant terms with variants in all the 11 languages

WEBSTUDIO: collaborative work of BMS experts from all over the world

The screenshot displays the WEBSTUDIO interface for managing a multilingual thesaurus. On the left, a tree view of hierarchy shows a list of terms under 'materials', with 'coatings' selected. The main area shows a table of terms with columns for NAME, VARIANT, BROADER, VARIANT.FRENCH, and LANGUAGE. Below the table, a form for adding/editing terms is visible, with fields for 'BROADER' and 'VARIANT'. The 'BROADER' field contains 'materials' and the 'VARIANT' field contains 'revêtement de'.

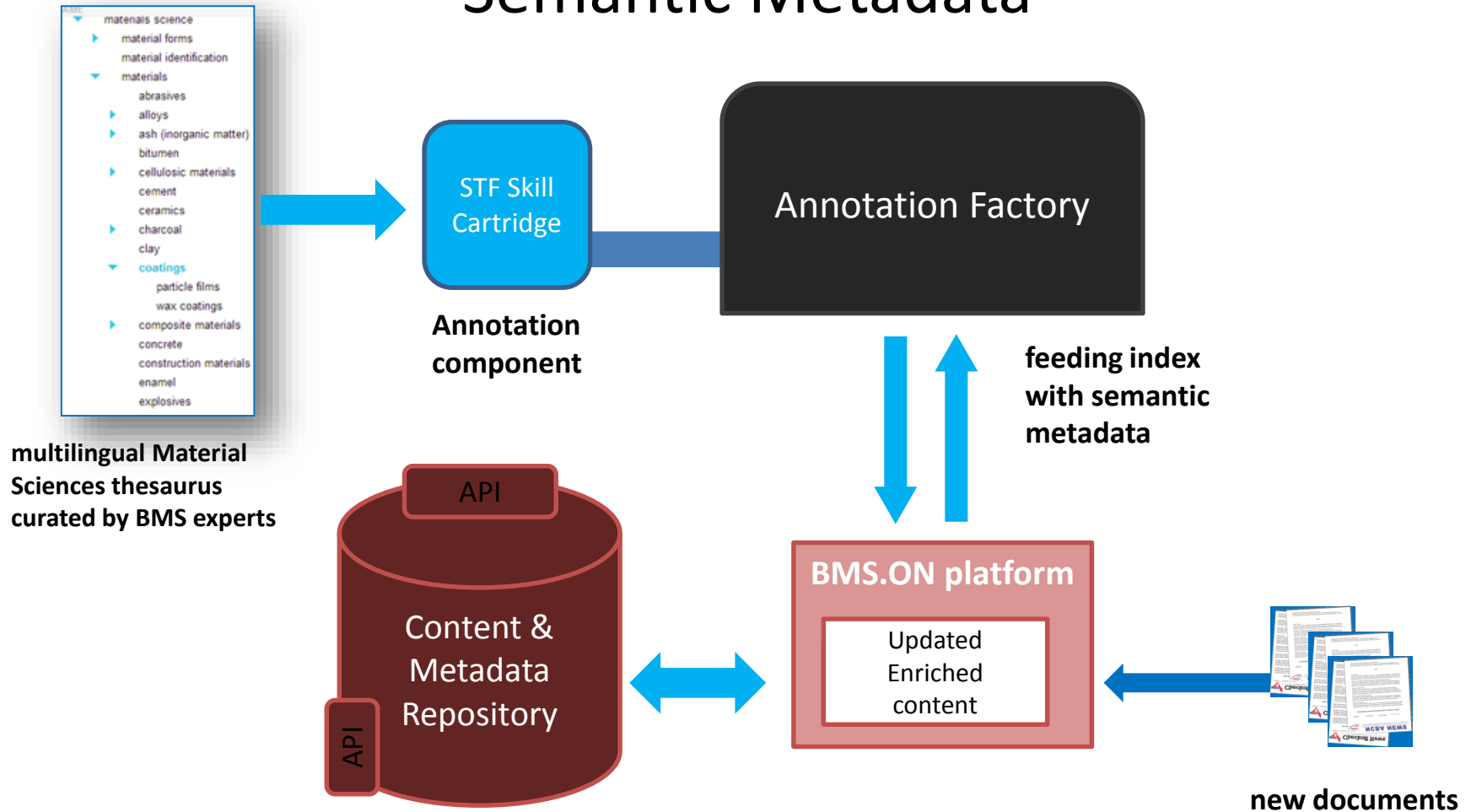
NAME	VARIANT	BROADER	VARIANT.FRENCH	LANGUAGE
particle films	HPF technology	coatings		english
wax coatings		coatings		english

tree view of hierarchy

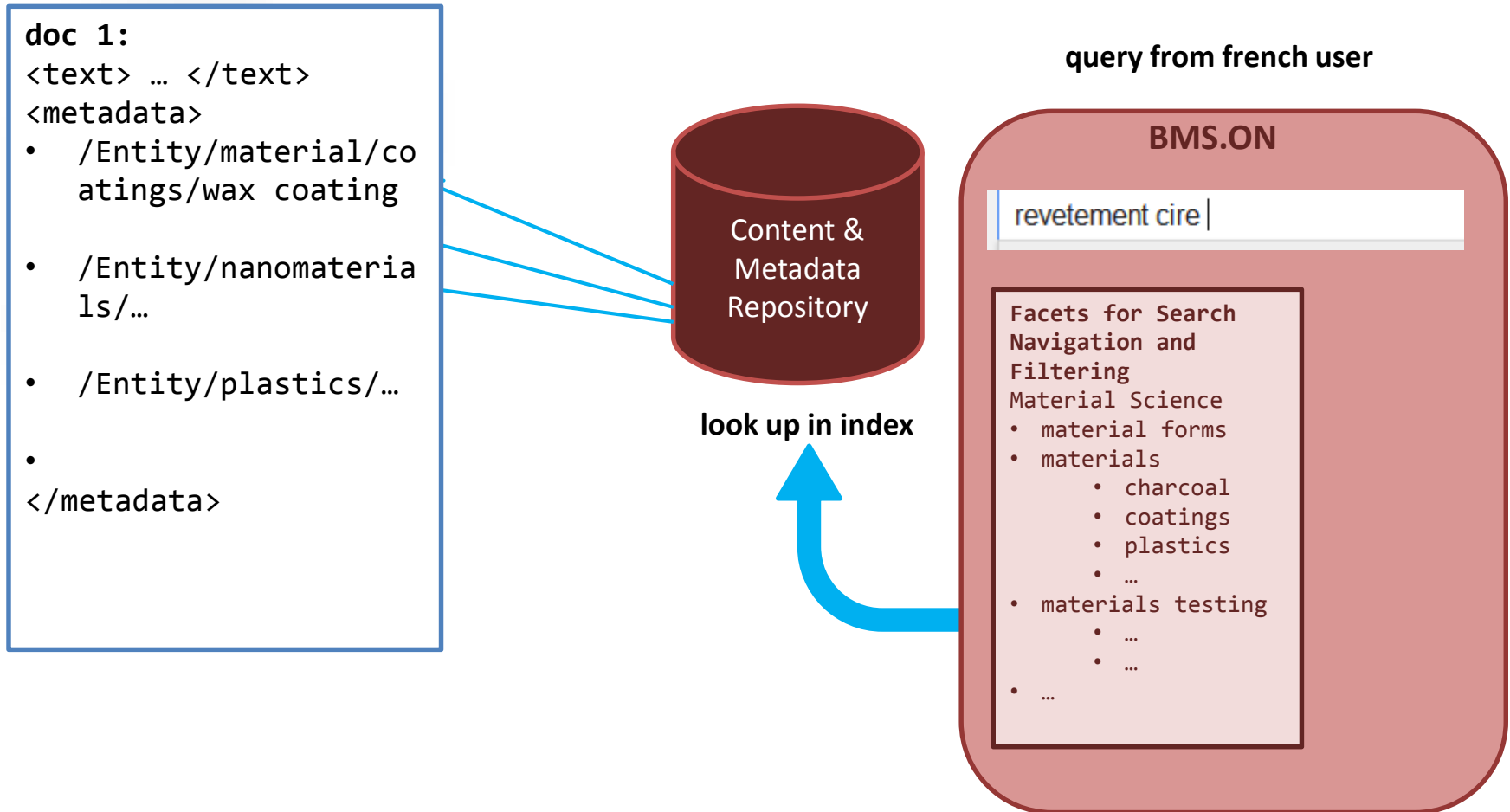
organize hierarchy

add / edit new terms and variants

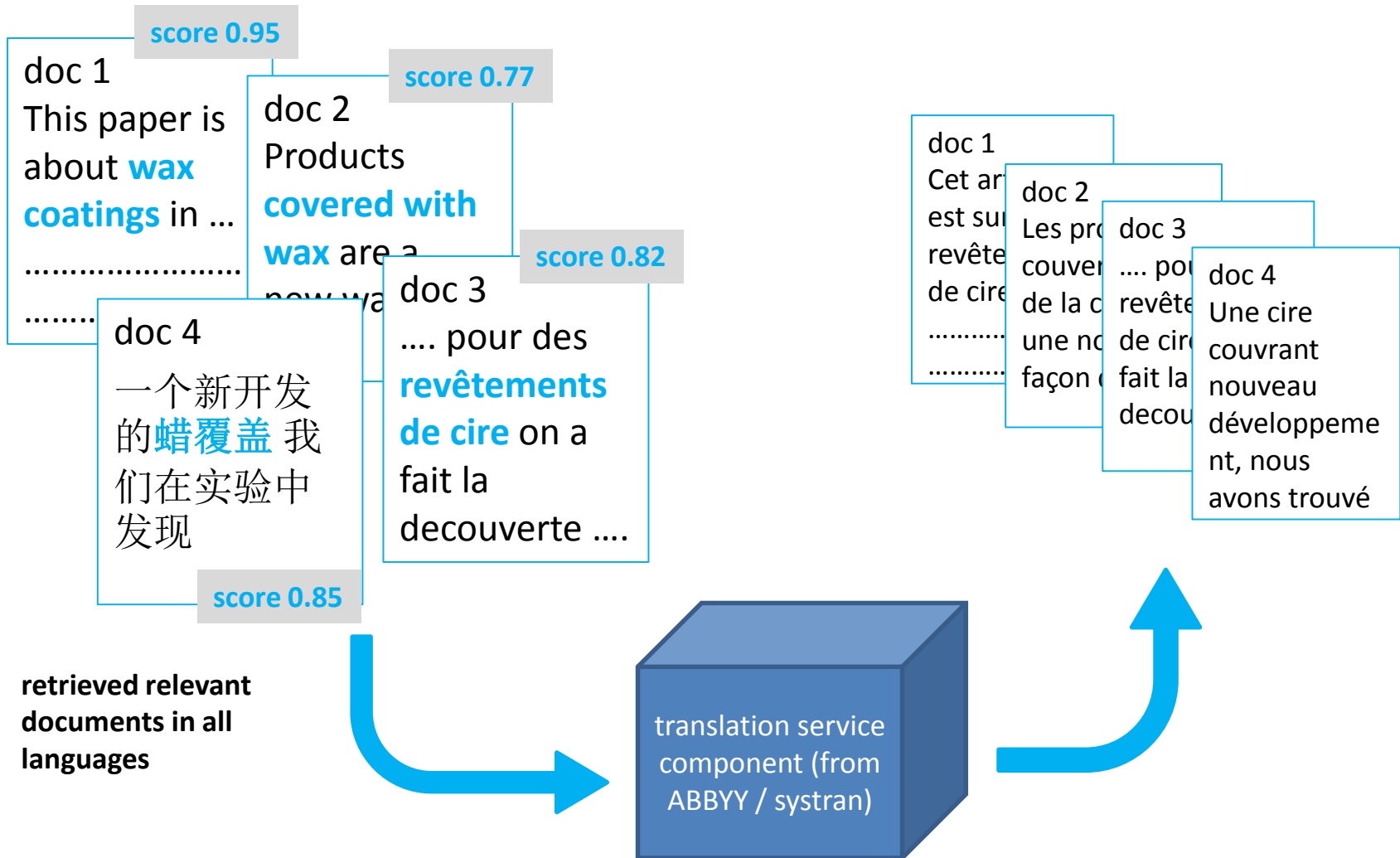
Annotation: Content Enrichment with Semantic Metadata



Search over semantic metadata



Search results



Imooty.eu GmbH & Co KG

Imooty's specialty is big data analytics i.e. to process text data from multiple sources, including web services, social media networks and databases.

The content from these sources is generally first downloaded by our crawling software, employing the appropriate APIs. As a second step, our computational linguistic algorithms classify the content, sorting it according to thematic categories, similarity, language and/or sentiment.

Accordingly, we could contribute three types of expertise for this project:

- A Duplicate Filter detects documents which have equal (or very similar) content
- A Language Detector tags input documents with one of 50 languages
- A Document Classifier assigns documents to pre-defined themes and topics

All our analytical tools are optimized for multilingual content. In combination with the subsequent translation process, they enable the user to find the most relevant documents for his scientific research.

References in the field of linguistic “Big-Data“ services are:

- Cortal Consorts, BNP Paribas subsidiary in Germany
- Montel, leading information portal for Europe’s energy market
- Kinetick, global provider of price data for the investment community