

ABBYY Intelligent Solutions for Intelligent Enterprise

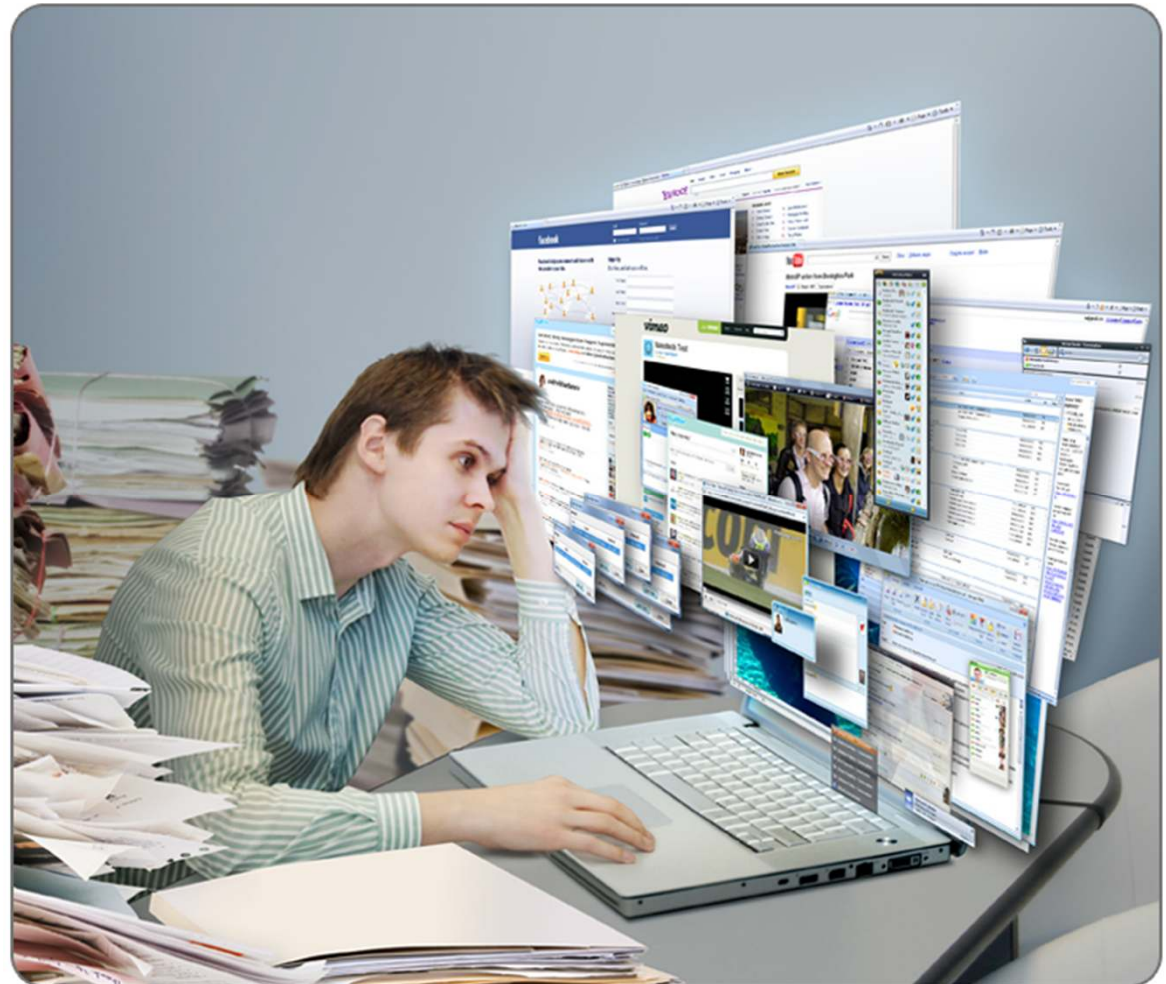


Alexander Rylov

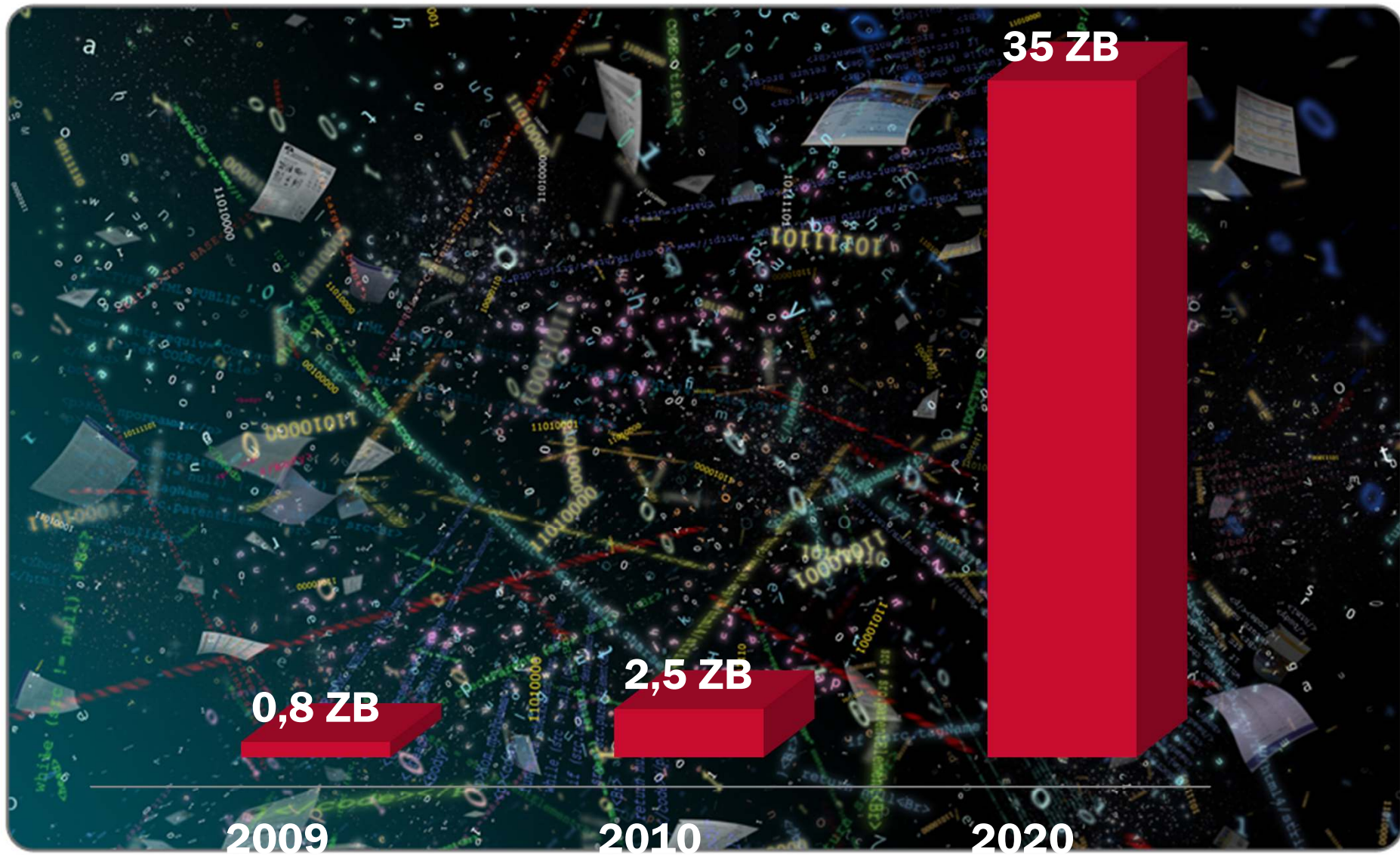
Director of Linguistic Products Department

Information explosion

- Digitization
- Computerization
- Internetization
- Globalization



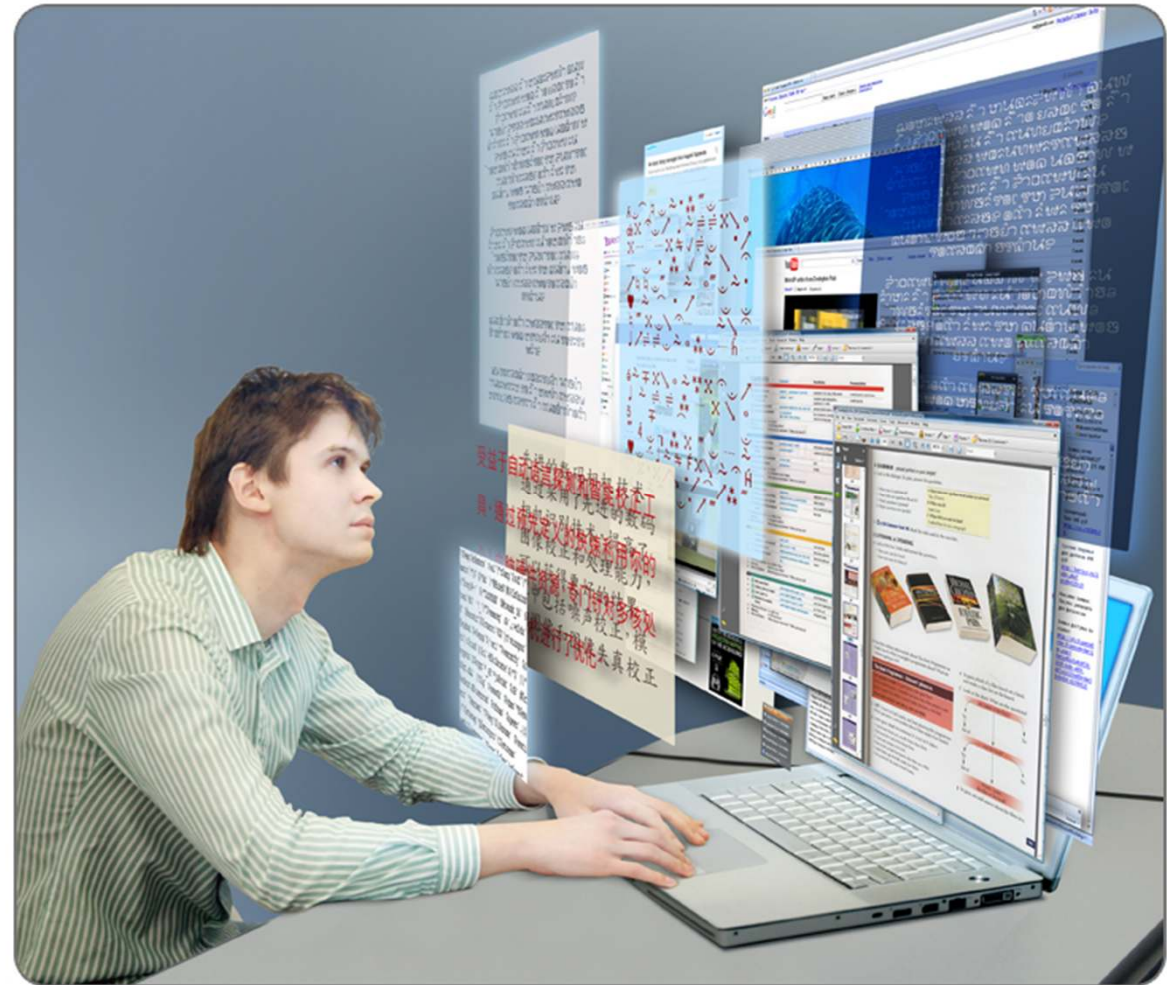
Digital Universe



По данным IDC, 05-2011

And even more than data amount...

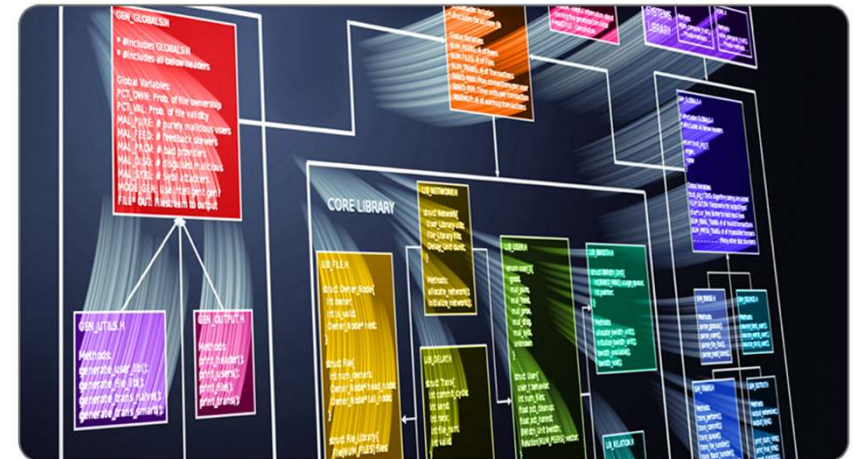
- Unstructured data
- Often not editable
- In different languages



ABBYY Technologies

Recognition (OCR)

- Data transforming (*paper/pdf – editable/MS-Office*)
- Data capturing from unstructured forms



Semantic and Syntactic text analysis

Building of semantic and syntactic representation of sentences

- Translation
- Smart search by meaning
- Facts and relations extraction

Challenge

«...

(Reuters) - **Business software company Yammer Inc has agreed to sell itself to Microsoft Corp for more than \$1 billion, the Wall Street Journal reported, citing a person familiar with the matter.**

It was not certain when the deal would be completed and formally announced, the Journal said.

San Francisco-based Yammer declined to comment, while Microsoft could not be reached for comment outside regular U.S. business hours.

Yammer, which began in 2008 and provides enterprise social networking services, has raised more than \$142 million in venture capital funding.

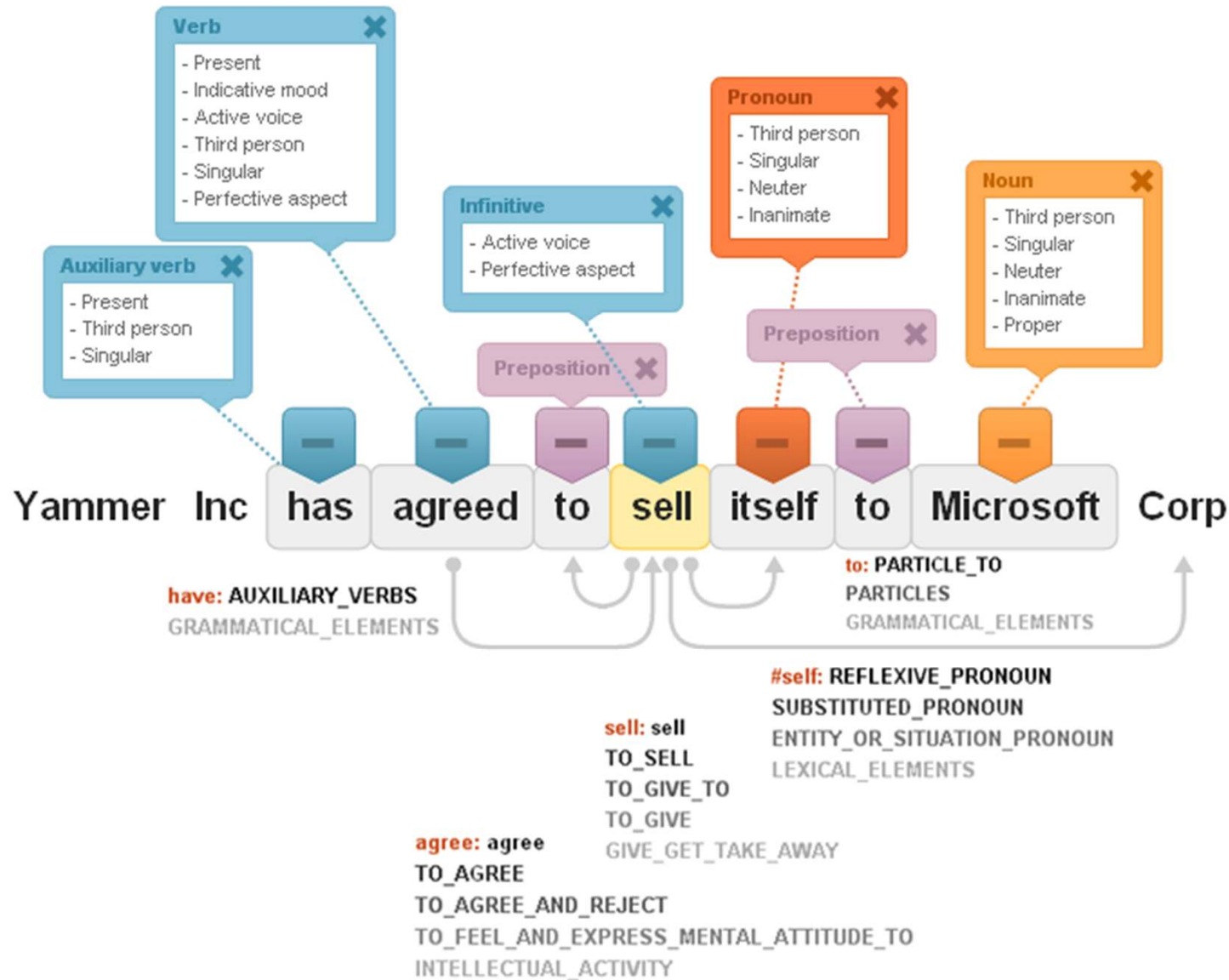
It says it counts more than 80 percent of Fortune 500 companies as clients, and is backed by PayPal co-founder Peter Thiel, who also invested in Facebook.

(Reporting by Jochelle Mendonca in Bangalore; Editing by [Ian Geoghegan](#))

...»

ABBYY Compreno

Syntactic and Semantic Text Analysis

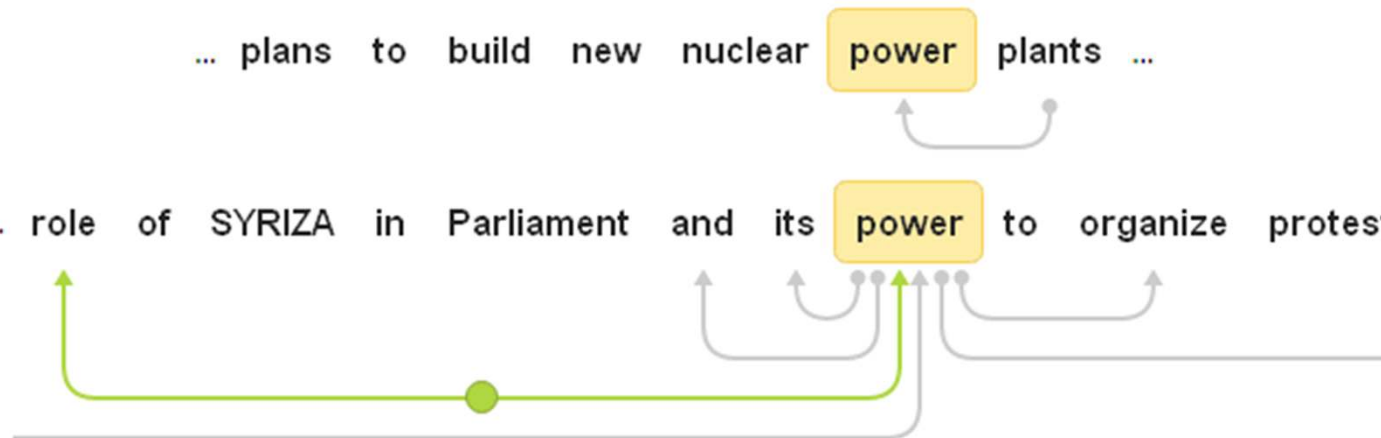


Meaning of words

power: POWER
 POWER_RATING_CHAR
 POWER_CHAR
 CH_PARAMETER_OF_OBJECT_AND_SUBSTANCE

... plans to build new nuclear **power** plants ...

... role of SYRIZA in Parliament and its **power** to organize protest ...

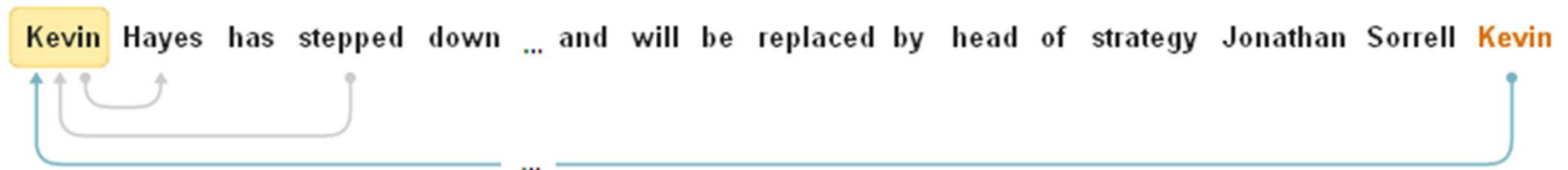


power: power
 POWER_AS_INTERNAL_ABILITY
 INTERNAL_ABILITY
 INTERNAL_EXTERNAL_ABILITY
 POSSIBILITIVE

Complex relations (Ellipsis)

«Kevin Hayes has stepped down as finance director of struggling hedge fund firm Man Group, and will be replaced by head of strategy Jonathan Sorrell.»

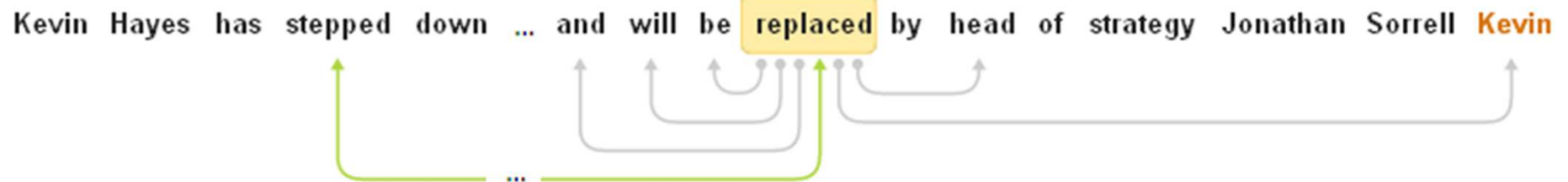
bbc.co.uk



Complex relations (Coordination, long distance)

«Kevin Hayes has stepped down as finance director of struggling hedge fund firm Man Group, and will be replaced by head of strategy Jonathan Sorrell.»

bbc.co.uk



Complex relations (Anaphora)



Ontological Objects

Business software company Yammer Inc has agreed to sell itself

to Microsoft Corp for more than \$1 billion,

the Wall Street Journal reported, citing a person familiar with the matter.

Relations



How can we change the search with ABBYY Compreno?

people win

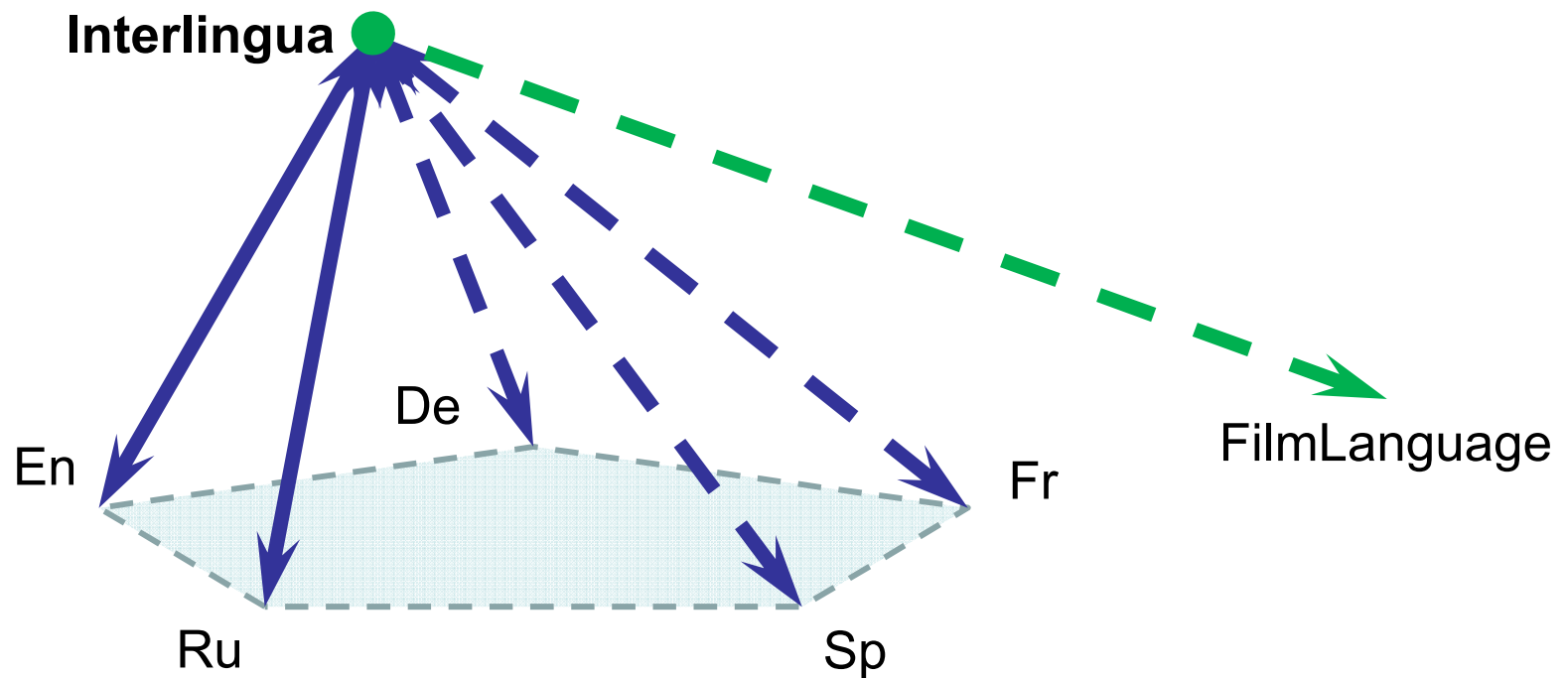


[Agent: "HUMAN"] "TO_WIN"

- When prices fall, **shorts gain** and longs lose.
- When a stock closes higher, it shows that **bulls won** the day's battle; its volume on that day is added to OBV.
- When a stock closes lower, it shows that **bears won** the day, and that day's volume is subtracted from OBV.
- When a **beginner wins**, he feels brilliant and invincible.
- **Traders gain** some knowledge, **they win**, their emotions kick in, and they self-destruct.
- A trader who feels overjoyed when **he wins** and depressed when he loses cannot accumulate equity because he is controlled by his emotions.

Basis of technology – Pure Interlingual approach

- Based on 50 years of fundamental linguistic science
 - Universal Semantic Hierarchy
 - Syntax parsing with the use of semantic
 - Statistically-based Verification and Adjustment on parallel and monolingual corpora



Multilingual search possibilities

где расположена база



[Locative:] "POSITION_IN_SPACE" [Object: "BASE_IN_BASEBALL"]

Сама база находится в Рендельшемском лесу.

В двенадцати милях от Ипсвича в густом лесу на побережье находится база Бентуотерс.

The whole base, it seems, is in the heart of Rendlesham Forest.

Twelve miles from Ipswich, in a dense forest and close to the shore, was Bentwaters.

В северо-восточной части промплощадки размещена база заказчика, имеющая в плане размеры 500x250 м общей площадью 12.5 га.

Размещение базы заказчика на начальном отрезке железнодорожной ветки позволяет обеспечить доставку на базу материалов и оборудования по железной дороге уже на первом этапе строительства.

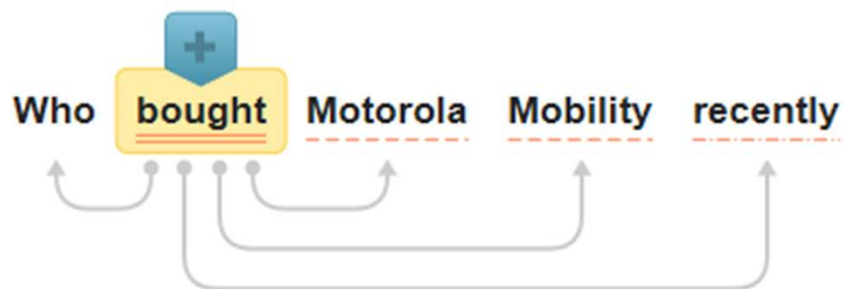
The Customer's base area having size of 500×250 m (total area 12.5 ha) is located in north-eastern part of the site.

Location of the Customer's base area on the first piece of railway branch allows the possibility of the delivery of material and equipment by railway, even in the first construction stage.

Multilingual search possibilities

Create on native language, perform search on all languages, provide result on native language

Search request:
Who bought Motorola Mobility recently?



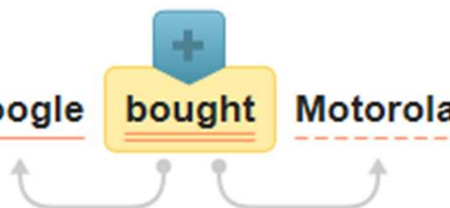
buy: buy
TO_BUY
TO_ACQUIRE
TO_GET
GIVE_GET_TAKE_AWAY

Search resource in Russian:
Google покупает Motorola за \$12,5 миллионов



купить: купить
TO_BUY
TO_ACQUIRE
TO_GET
GIVE_GET_TAKE_AWAY

Search answer: Google bought Motorola



Bazelevs - FilmLanguage

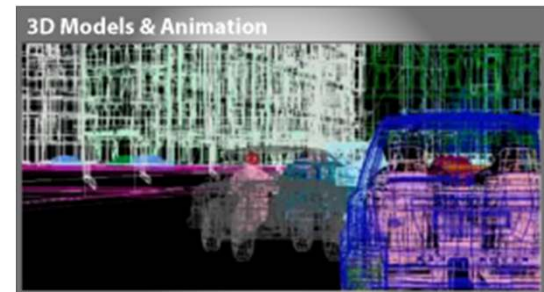
- Automatically translates natural text into 3D animation
- The comprehensive and accurate semantic and syntactic parser is key for visualizing texts

Text

John is driving through the city on his black retro car, while being chased by a red Range Rover. The cars swerve in between traffic, merging

Script

```
<ExtendedLexicalClass>
John:JOHN:PERSON_BY_
FIRSTNAME:HUMAN_BY_
NAME:BEING_BY_NAME
</ExtendedLexicalClass>
```



Most important applications

- Voice of customer (Sentiment analysis)
- Information monitoring
- Business intelligence
- eDiscovery
- Semantic multilingual search
- Machine Translation

Ways for collaboration

- For products vendors and integrators
 - Include ABBYY Compreno into your existing solution and improve quality of search and text analysis
 - Make mutual search and text analysis solutions
- For technology vendors and linguistic scientists
 - Add descriptions to new languages in ABBYY Compreno
- For universities and academics
 - Increase in the number of linguistic specialists, creation of jobs, benefits to science.

Few words about ABBYY

One of the largest global research centers in linguistics

- Founded in 1989
- 14 offices worldwide
- More than 1200 employees
- About 560 employees in R&D (Russia)
- Licensing ABBYY technology to more than 1000 partners worldwide
- Dept. of computer linguistic science in 2 universities: MIPT and RSHU in Moscow
- ABBYY Compreno
 - 17 years
 - 1700 men-years
 - About 350 linguists, programmers and analysts involved in ABBYY Compreno
 - Patented technology in US
- Prototypes for
 - Fact Extractor
 - Semantic Search
 - Machine Translation



Thank you for your attention!