Author: Laura Tolosi

What better to do on a rainy working Saturday in Sofia than a Hackathon? Teams are ready, ideas are ambitious and pizzas are hardly enough for the high-calorie consumption that thinking out of the box requires.

Needless to say, we learned a lot. We teamed with colleagues we didn't work with closely before. We were goal-oriented and pragmatic enough and, for a day and a half, we managed to deliver a meaningful result, convincing the public that there was real value in the prototypes developed.

And maybe the most valuable takeaway was that Ontotext technologies are not rigid or limited to a set of commercial products, but can be exploited in new, innovative ways. Expect to see some of those put into practice soon!

Here's an overview of the projects:

(Semi-)automated semantic market analysis

Traditional market analysis is a tedious process that is customized and carried out manually for each beneficiary. Currently, keyword analysis of competitive websites determines the key concepts around which the competition is trying to position themselves on the online market. Research analysts in a company gather information relevant to the market from a variety of resources, including government agencies, primary and secondary research.

Semantic Technology can automate much of the workflow of market research. Given a set of resources pointed as relevant to the company, our tools can extract the entities that denote the big players in the industry, their relations, dynamics in the news, etc. This was a non-technical project but it provided a good use case for a platform like FactForge-News, used in the Today's News project (see below).

Social Media Monitor

Ontotext's focus on Social Media analysis has been secondary to the traditional publishing channels. We are involved in the Pheme project, which gives the premises for future development of Social Media analysis. This small toy-project involved the analysis of our own Twitter accounts, which revealed the main topics that Ontotext cares and speaks about.

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

ontotext

Stop accidental consumption of banned substances

Every year, each sports updates and publishes a list of banned substances against doping. Ontotext's Linked Life Data (LLD) can be used for automated recognition of banned substances in text, may that be a list of ingredients, a list of active substances in a drug, etc.

The team managed to augment LLD with additional data from PubChem and successfully implemented such a checker during the hackathon. As a test, the ITF prohibited drug list was normalized to 11 955 distinct compounds with an overall of 97 330 literals. The implemented pipeline was able to identify both meldonium and mildronate, which were recently added to the ITF banned drugs list, and because of which Maria Sharapova was suspended.

Word vectors to improve text analysis models

Word vectors are the promising first steps of deep learning in the field of Natural Language Processing. We want to take Ontotext's text analysis tools to a higher level of performance by implementing word vectors features and deep neural networks to serve document classification (including sentiment analysis, topic classification, keyword assignment), Named Entity Recognition, document clustering, entity clustering, topic modeling, content recommendation, etc.

At the end of the day, we improved the current models for multi-label classification by 4%! And they were pretty good to start with, showing more than 80% F1 on a challenging use case that we developed for The IET. We also managed to cluster named entities occurring in interviews with Holocaust victims (we are part of the EHRI project), which should give us a start for discovering rare spellings of locations, people and organizations involved in the Holocaust.

Today's News Map

Ontotext's news showcase NOW.ontotext.com has been accumulating semantically annotated news for more than a year, linking them to a rich collection of Linked Open Data like DBpedia, Geonames, etc. Think of 120 thousand news, linked with 7 million tags to a Knowledge Graph of 500 million statements, describing more than 7 million entities and resources.

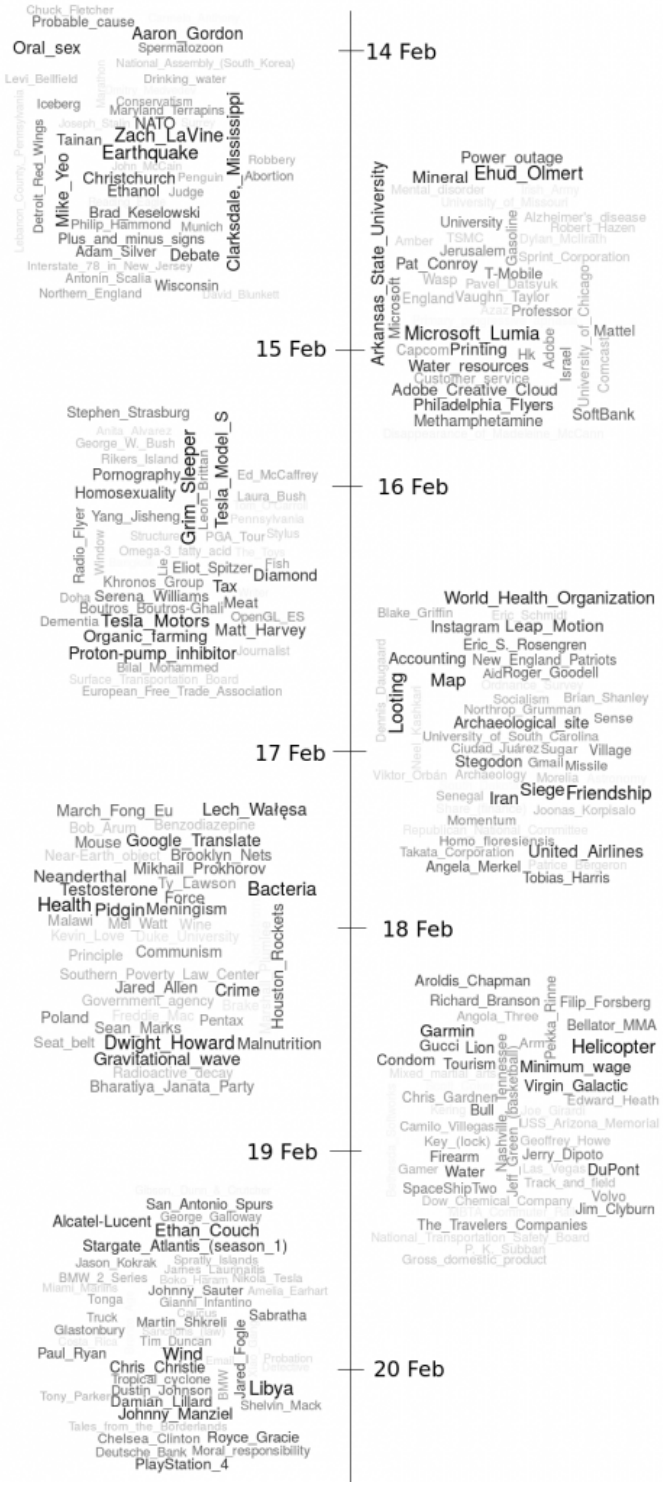
Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

On average, there are 70 annotations of news with identifiers of specific concepts in the Knowledge Graph. The integrated resources reveal subtle and rich information, suffice it to know SPARQL well enough! And if you don't, it's definitely worth learning!

With a SPARQL query and some normalization of frequencies using z-scores, one can find the most relevant entity mentions for the day in the news. Here's what's going on in the news between Feb 14th and Feb 20th, 2016 (the picture below). With a proper UI, all these entities would be clickable and leading to news and DBpedia articles.

This platform will be released for public access and experimentation during the March 24 webinar "Boost Your Data Analytics with Open Data and Public News Content".

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com



Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

ontotext

SPARQL Tree Matcher

We have been writing SPARQL queries for a long time in Ontotext. It's about time we analyze them, too. We love the recursive logic here: let's data mine our data mining tools!

GraphDB gives us access to log files that represent submitted queries and their parse trees. We wanted to visualize the population of queries and to cluster them. So we used a trick: we represented SPARQL statements as protein amino acids in order to take advantage of so many algorithms for computing phylogenetic trees.

So, we had our own population of queries, with their sisters and cousins and distant relatives. On a practical note, this kind of tooling can help our support team to faster analyze the logs of our enterprise clients' production systems and to be much more efficient in identifying problematic query patterns.

BG NOW

Ontotext's semantic analysis and search technologies are in a process of diversification. One important direction is to offer multi-language support. We already have solid experience with German, Italian, Dutch, Bulgarian and French.

This project took the necessary steps towards publishing Ontotext's NOW.ontotext.com semantic news showcase in Bulgarian. So far, it only contains news from one big Bulgarian online publisher (a client of ours) and it is not public. Here's an early preview:

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

News

Наука и технологии

17 Oct 2015
ЕС и Русия изпращат общ изследователски апарат до Луната (видео)
Европейският съюз и Русия планират до 5 години да изпратят заедно космически изследователски апарат до Луната, съобщи BBC News. Първата роботизирана

17 Oct 2015
Осъдиха Apple да плати над $230 млн. заради нарушен патент
Съдът в Уисконсин реши, че световноизвестната технологична компания Apple трябва да плати на местния университет сумата от 234 млн. долара заради

15 Oct 2015
Импулсна пушка укротява нежелани дронове (видео)
Някои собственици на къщи стрелят с пушки по нежеланите над техните имоти дронове. За полицията този подход не е подходящ, а пък й липсва

15 Oct 2015
Как да рекламираме успешно във Facebook
Един от най-важните критерии, с който се измерва успеха на една реклама в дигиталния маркетинг, са т.нар. конверсии (коефициентът, който показва

Анализи и коментари

17 Oct 2015
Ксенофобия и хиперболизиран либерализъм след инцидента на границата
Инцидентът, при който загина афганистански имигрант край град Средец, се превърна в поредното доказателство за дълбокото разделение и

17 Oct 2015
Здравка Евтимова: Октомври
Чувах я преди години в мислите си, тиха мелодия, съвсем обикновена, и си представях път към малък град. Знаех, че ще си тръгнеш. Песента идваше като

16 Oct 2015
Охраната на границата като "рикошет от въздуха"
Въпрос на време беше да се стигне до инцидента в близост до границата, при който беше убит нелегален имигрант . За нечовешките условия, при които

15 Oct 2015
„Реставрация“, „консервация“ или ново строителство: бутафория и заслужена жлъч
Пиша този текст по повод материала „Реставрацията на Ларгото: бутафория ли пресилена жлъч“ от г-н Калоян Константинов. Изглежда, че е нужно

Живот

17 Oct 2015
Св. прор. Осия, Св. прпмчк Андрей Критски
Свети пророк Осия, син Беериев , произхожал от Исахаровото коляно (1). В онези времена много израилтяни отстъпили от истинския Бог и като забравили

16 Oct 2015
Дядо Коледа стана общински съветник в Аляска
Отклонявайки погледа от кампанията за местните избори на 25 октомври в България ви представяме Дядо Коледа, който вече стана общински съветник в

16 Oct 2015
Бенисио дел Торо ще е злодеят в осмата част на Междузвездни войни
Бенисио дел Торо ще бъде главният злодей в предстоящата осма част на космическата сага Междузвездни войни, която в момента се снима и трябва да

16 Oct 2015
Над хиляда души завлечени с 60 000 лв. от провален концерт
Над 1000 души изгубиха парите, които са заплатили за билет за проваления концерт на американската певица Лорин Хил, предаде "Нова телевизия".

Спорт

17 Oct 2015
ВК Левски и Владо Николов правят волейболно училище в София
Капитанът на волейболния ни национален отбор и президент на ВК Левски Владимир Николов ще основне волейболно училище съвместно с клуба. Школата ще

17 Oct 2015
Билети за България - Франция все още има
Билети за полуфинала на България на Евроволей 2015 все още има. На касите на "Арена Армеец" и в момента има опашка от хора, чакащи да си закупят

17 Oct 2015
Багата защитава Световната си титла тази нощ в САЩ
Благой Иванов ще защитава за първи път Световната си титла в смесените бойни изкуства на WSOF в тежка категория тази нощ. Българинът спечели трофея

17 Oct 2015
Кубрат Пулев се връща на ринга срещу Джордж Ариас тази вечер в Карлсруе
Кубрат Пулев се завръща на професионалния ринг тази вечер след почти едногодишно отсъствие след загубения мач за световната титла срещу Владимир

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

ontotext

17 Oct 2015

ЕС и Русия изпращат общ изследователски апарат до Луната (видео)

Европейският съюз и Русия планират до 5 години да изпратят заедно космически изследователски апарат до Луната, съобщи BBC News. Първата роботизирана съвместна мисия носи името "Луна-27" и трябва да разбере дали има вода и суровини за синтезиране на горива и кислород. Апаратът ще кацне в басейна Айткен, в близост до южния полюс на Луната. Южната полярна област има райони, които са постоянно тъмни и е не само едно от най-мразовитите места в Слънчевата система, но и една от най-слабо познатите й части. Учените от "Роскосмос" и Европейската космическа агенция се надяват, че ще открият ресурси, които ще позволят извличане на ракетно гориво и поддръжка на животоподдържащи системи за жителите на бъдеща постоянна колония на Луната. Според д-р Джеймс Карпентър (James Carpenter), водеш учен по проекта на ЕКА, една от основните цели е да се проучи потенциалното използване на водата като ресурс за бъдещето, и да се разбере какво може да ни разкаже за произхода на живота във вътрешната част на Слънчевата система. В съвместните планове е включено и изграждането на международна постоянна база на обратната стана на Луната. До края на 2016 г. ЕС трябва да реши напълно дали ще участва, като засега е дала предварително съгласие. Но според научният редактор на BBC News Дейвид Шъкман (David Shukman) следващият отпечатък в лунния прах ще бъде китайски. Китай има амбициозни планове да експлоатира лунни мини за извличане на ценни ресурси като хелий-3. През цялата история на човечеството на Луната се гледаше по различен начин. През 60-те, тя бе сцена на съперничеството от Студената война. Сега на нея се гледа като на потенциална междинна станция за по-дълги космически пътешествия и като на мина, която чака да бъде изкопана и експлоатирана.

Related

16 Jan 2015

Петте най-луди космически мисии на бъдещето

BBC предлага описание на петте най-невероятни експедиции, които се надяват да осъществят

16 Dec 2014

Кои са значимите постижения в космоса за 2014

След дълги години на затишие, 2014-та бе най-вълнуващата година за космическите изследвания

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

So, the road ahead is full of exciting new opportunities to prove the value of embracing cognitive and semantic technology!

Ready to join us?

Contact us for a consultation

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com