ayfie

WORDS MATTER

Introduction

Humans continuously curate knowledge in the form they know best: natural language. As most enterprise content became digitized, this created a flood of unstructured information, now often referred to as big content or the unstructured side of big data.

The mountains of unstructured enterprise content is available as natural language text, accessible to humans who are willing to invest the time to read it, but usually inaccessible to all but the most rudimentary automated analyses.

Because as much as eighty percent of an organization's total information assets are unstructured, and they are mostly unexplored, this has become the largest source of untapped potential insight being neglected in the big data revolution.

Examining Wikipedia and its structured siblings DBPedia and Wikidata, it becomes obvious that much more information is generated in unstructured form today than is codified in formats accessible to computation. The stalled momentum of the semantic web movement painfully illustrates this.

ayfie makes that trove accessible to analysis and prediction algorithms by unearthing the sematic structure behind the text.

Johannes Stiehler
Chief Technology Officer, ayfie Group AS

The key to ayfie's intuitive and insightful information retrieval results is its understanding of linguistics and its ability connect natural phrases across any corpus of documents.

Ingestion

Using ayfie, documents can be collected from almost any common storage system or application, from cloud storage providers to ECM tools to information stored in CRM systems and historical databases.

Of course, documents can also be pushed through the ayfie API in order to facilitate integration into 3rd party applications.

After collection, documents are analyzed, enriched, and converted into a common text representation, keeping all meta-data and structural information intact.

ayfie benefits from on a decade of experience in the enterprise search market where connecting to many different systems – from content management to email servers – is key to success. Over 70 connectors are available and can be extended using a robust SDK.
Text and metadata can be extracted from all popular document formats. Text contained in images is extracted using OCR technology.
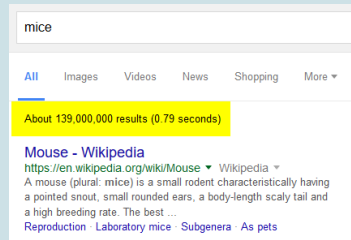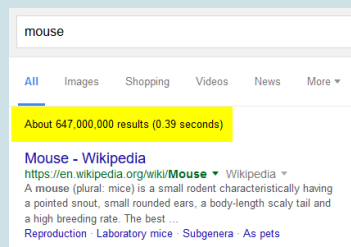
Enrichment

By applying large electronic dictionaries that codify information about semantic and syntactic properties of words (single and multi-term expressions), ayfie enriches texts with base forms, synonyms, known spelling variants and dependency properties for all words occurring in the text.

It finds and annotates all external structure that exists in the documents such as headings, paragraphs, sentences etc. and makes them accessible to further analysis and examination.

ayfie builds on more than 30 years of research in compiling large scale electronic dictionaries and other linguistic resources. Inflectional forms, synonyms and other phenomena (such as decomposition in Norwegian and German) are handled for all major European languages.

Word inflections continuously lead to inconsistent search results on Google.

This effect is of course multiplied for multi-term queries.

For dealing with structure extraction and semantics, ayfie builds on well-researched linguistic frameworks based on the works of Zelig Harris, Maurice Gross and Franz Guenthner. This sound theoretical foundation is combined with our blazingly fast proprietary extraction engine and exhaustive resources in many languages and domains.

Many of our customers actually use this functionality alone in order to greatly improve the search experience on their customer-facing portals.

Dealing with these linguistic properties of words and phrases is key to consistent and relevant search results, especially in domain-specific applications.

Structurization

ayfie analyzes the semantic structure inside the document insofar as it is relevant to the use case and brings it into a form where it is more efficiently consumable for humans and formal enough for machines.
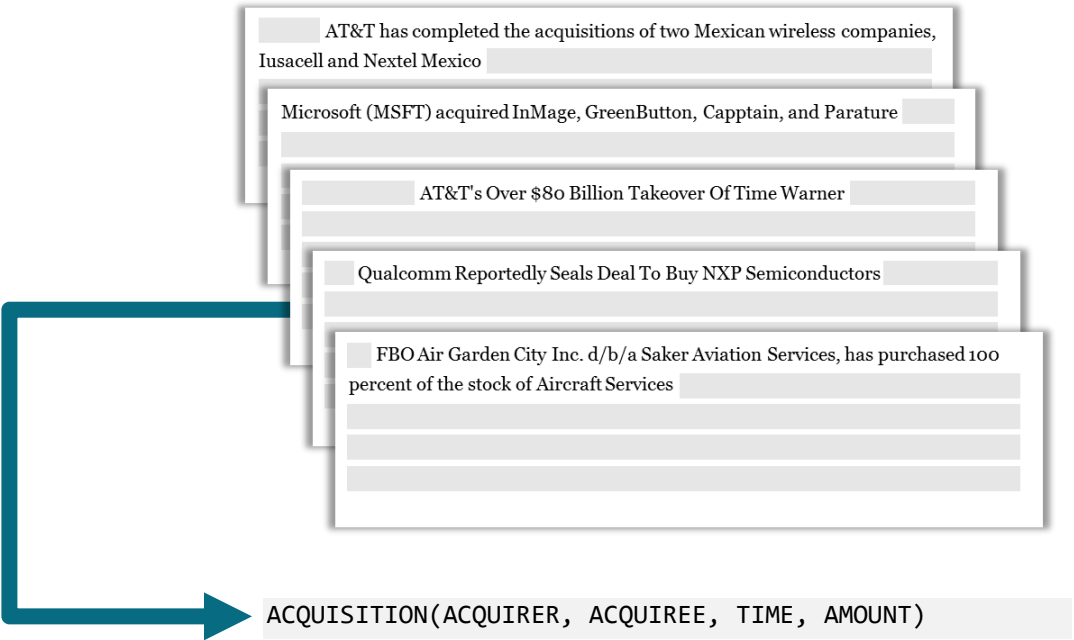
"What distinguishes ayfie from all previous and current approaches to text analysis is our view of the basic elements of meaning in language. Our algorithms do not simply manipulate isolated words - which are always either strongly ambiguous or extremely vague - but complex semantic constructions which express meanings at a higher level.

We apply combinations of very large semantic dictionaries that encode a lot of information about the entities mentioned in the texts together with millions of semantically typed parsing rules that know how relations are expressed syntactically.

We are therefore able to extract the names of entities, the facts and the opinions expressed in the text."
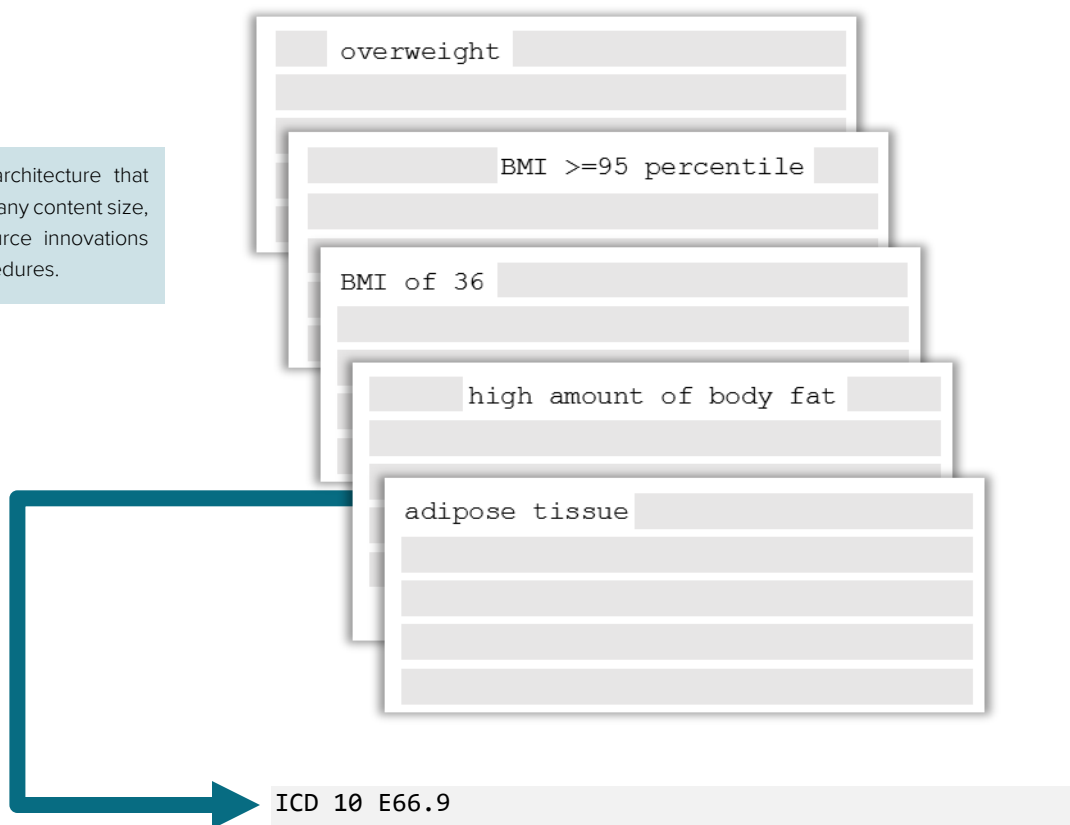
Franz Guenthner, professor of Computational Linguistics at the Center for Information and Language Processing at the Ludwig-Maximilian-University (LMU) and Technology Advisor at ayfie Inc.

For instance, there are thousands of ways to describe the acquisition of one company by another. ayfie recognizes them all and makes them accessible to search applications and further statistical analysis.

AT&T has completed the acquisitions of two Mexican wireless companies, Iusacell and Nextel Mexico

Microsoft (MSFT) acquired InMage, GreenButton, Capptain, and Parature

AT&T's Over $80 Billion Takeover Of Time Warner

Qualcomm Reportedly Seals Deal To Buy NXP Semiconductors

FBO Air Garden City Inc. d/b/a Saker Aviation Services, has purchased 100 percent of the stock of Aircraft Services

```
ACQUISITION(ACQUIRER, ACQUIREE, TIME, AMOUNT)
```

In the same manner, doctors use many different expressions, ranging from colloquial to very formal, in describing obesity in patient notes.
ayfie understands all of these variations and can map them to the correct ICD 10 code.

```
          overweight

                    BMI >=95 percentile

   BMI of 36

              high amount of body fat

      adipose tissue
```

```
ICD 10 E66.9
```

Storage

All extracted information is stored in the context of the original documents. This structured linguistic meta-information facilitates research, further processing and analysis. ayfie's storage architecture is scalable from small installations on a single consumer notebook, indexing e.g., tens of thousands documents for an eDiscovery case, to an arbitrary number of nodes storing all scientific articles of the biggest publishers in the world.

ayfie supports both graph-based and search-based access methods to efficiently execute many different types of algorithms on the extracted information and the raw data. A vast set of operations is included out of the box. Additional ones can be added in the form of Spark jobs or ElasticSearch plugins, accessing not only the text, but the results of all prior structurization steps.

Applications

By extracting the most important information from any kind of document or corpus, ayfie powers a wide variety of applications in several different domains.

ayfie can automatically extract persons, locations, organizations, key phrases and many more entities out of the box that are perfect "hand rails" into the content to be searched.

Suggest

While it is relatively simple to construct a decent suggest / type-ahead search functionality from structured data such as an ecommerce product catalogue, the same is not true for unstructured data such as news text.

ayfie can for instance extract the most important concepts, person names, locations and organizations from financial news so they can be displayed on "Page Zero" of your application, as Microsoft calls it:

c	🔍

🏷 **persons**
carl icahn
steven cohen
charles plosser
charles evans
🏷 **locations**
china
🏷 **keywords**
crude oil price
crude oil
cash flow
capital expenditure
rig count

⌄ Show more

Search

ayfie's linguistic preprocessing and extraction capabilities make search applications much more consistent and exhaustive – but without sacrificing relevancy due to tricky phenomena like synonyms, inflectional forms and other variations.

ayfie's advanced extraction engine can even be used to power natural language search by reducing both the query and the document content to their semantic core.

Thus, the query

> Who acquired Aircraft Services Inc.? 🔍

matches the document content

> FBO Air Garden City Inc. d/b/a Saker Aviation Services has purchased 100 percent of the stock of Aircraft Services.

when processed by ayfie, because both sentences are reduced to a meaning representation of

```
ACQUISITION(FBO Air Garden City Inc. d/b/a Saker
Aviation Services, Aircraft Services Inc., ?, ?)
```

and

```
ACQUISITION(X, Aircraft Services Inc., ?, ?)
```

and are thus considered a match of each other.

Aggregation

By aggregating extracted predicates of the same type, ayfie can build tables of events or facts in any domain required. For instance, a large industrial manufacturer uses ayfie to analyze incident reports about turbine malfunctions. While these reports are written in plain text by service personnel, ayfie is able to turn them into tabular structures:

Ring	Position	Finding	Recommendation
Xla2	16	Crack	Replaced
Xla3	7	Wear	Smoothing
…	…	…	…

Based on the tables created by ayfie, the manufacturing company can now precisely answer questions like

> *"Which part failed most often because of outside heat?"*

> *"Which malfunction is the most likely for a certain class of parts?"*

> *"Which environmental conditions are the most detrimental to the overall reliability of the component?"*

Thus, all free-text service reports and even historical texts can be subjected to rigid analyses that would normally only be possible on pre-structured data.

Guided Machine Learning

Pure statistical and machine learning approaches promise to yield good results with any kind of text once the initial work of implementing the algorithm is done. However, training these algorithms inevitably requires training data, and lots of it if the results are to be useful.

> ayfie comes with a wide range of visualization and analysis options and can even drive 3rdparty data analysis tools.

The more structure we can give to text by breaking it into tokens, finding the boundaries of sentences, dealing with synonyms and inflectional forms, extracting salient phrases and entities, the closer we get to a structured representation of the text.

Thanks to this structure, the data can be processed more easily by learning algorithms that look for patterns and trends in the data. For instance, clustering and categorization yield higher precision when ayfie's structured representation of the text is used instead of simple token vectors. This produces a higher gain in precision than using a better learning algorithm on the unstructured text would.

Where ayfie is heading

At its core, ayfie is a technological platform capable of powering many different applications in a variety of markets. Currently, we are focusing on eDiscovery and Compliance in text-heavy and regulated industries including Legal, Insurance, Finance, and Healthcare.

Contact

For more information please do not hesitate to contact us at the addresses given below:

Rob Wescott · rob.wescott@ayfie.com
Chief Strategy Officer, ayfie Inc.
www.ayfie.com