

Myths and reality of text analytics

Andrew Marritt | Founder & CEO

andrew.marritt@organizationview.com | +41 78 905 0775

Organization View.

Our company. Started as one of the first People Analytics practices before pivoting to a product business in 2016

www.organizationview.com

Workometry.

By OrganizationView

Our product. Addresses the need to understand employee perception data provided as multi-lingual text. Now the main part of our business

www.workometry.com

Workometry success at innovation / startup competitions



Recognised by Gartner as a “Cool Vendor” in Human Capital Management Applying AI, 2018

What analysts are saying about applying AI in Human Capital Management.



“AI has made its way across the various HCM workstreams to have an impact on just about every HCM outcome. The Cool Vendors highlighted here offer solutions aimed at innovating employee engagement and retention, recruitment and experience, diversity and inclusion, and ultimately performance and efficiency. Whether selected by application leaders to drive better outcomes or analyze performance, these Cool Vendors are being highlighted specifically for their use of AI in helping to transform HCM and align technology with key objectives”

Source: “Cool Vendors in Human Capital Management Applying AI” by Jason A. Cerrato, Ron Hanscome, Helen Poitevin, Werner Goertz, John Kostoulas, Jeff Freyermuth 4 May 2018. The Gartner Cool Vendor Logo is a trademark and service mark of Gartner, Inc., and/or its affiliates, and is used herein with permission. All rights reserved. Gartner does not endorse any vendor, product or service depicted in its research publications, and does not advise technology users to select only those vendors with the highest ratings or other designation. Gartner research publications consist of the opinions of Gartner’s research organization and should not be construed as statements of fact. Gartner disclaims all warranties, expressed or implied, with respect to this research, including any warranties of merchantability or fitness for a particular purpose.

Ask questions about your pressing business issues

5

“How can we improve customer experience in our stores?”

“What could we do to improve the performance review process?”

“What could we do to improve your working life here?”

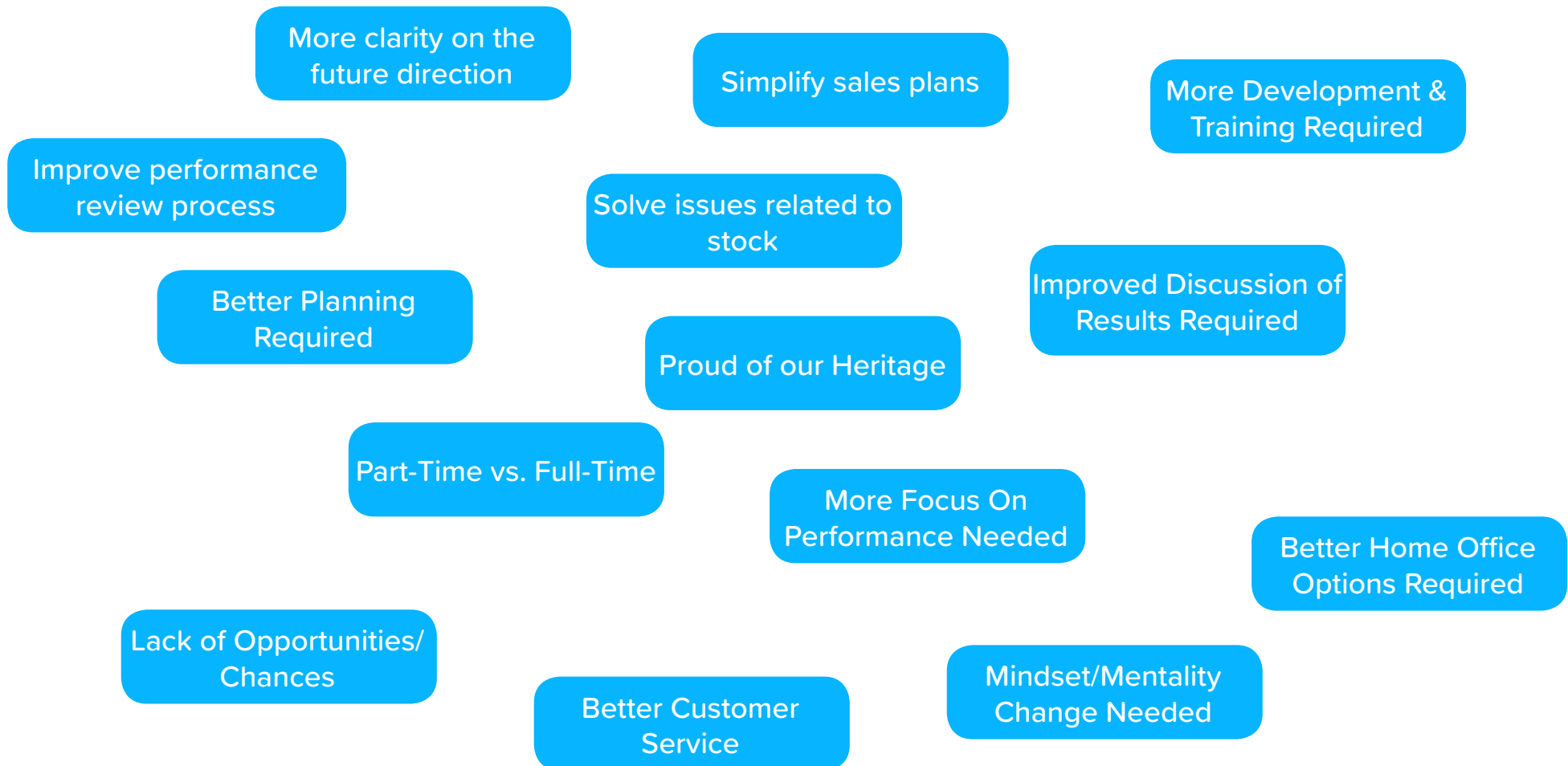
“What ideas do you have to improve the Christmas sales period?”

“How could HR improve the services they provide you as a manager?”

“If you were our Brand Director, what would you do first?”

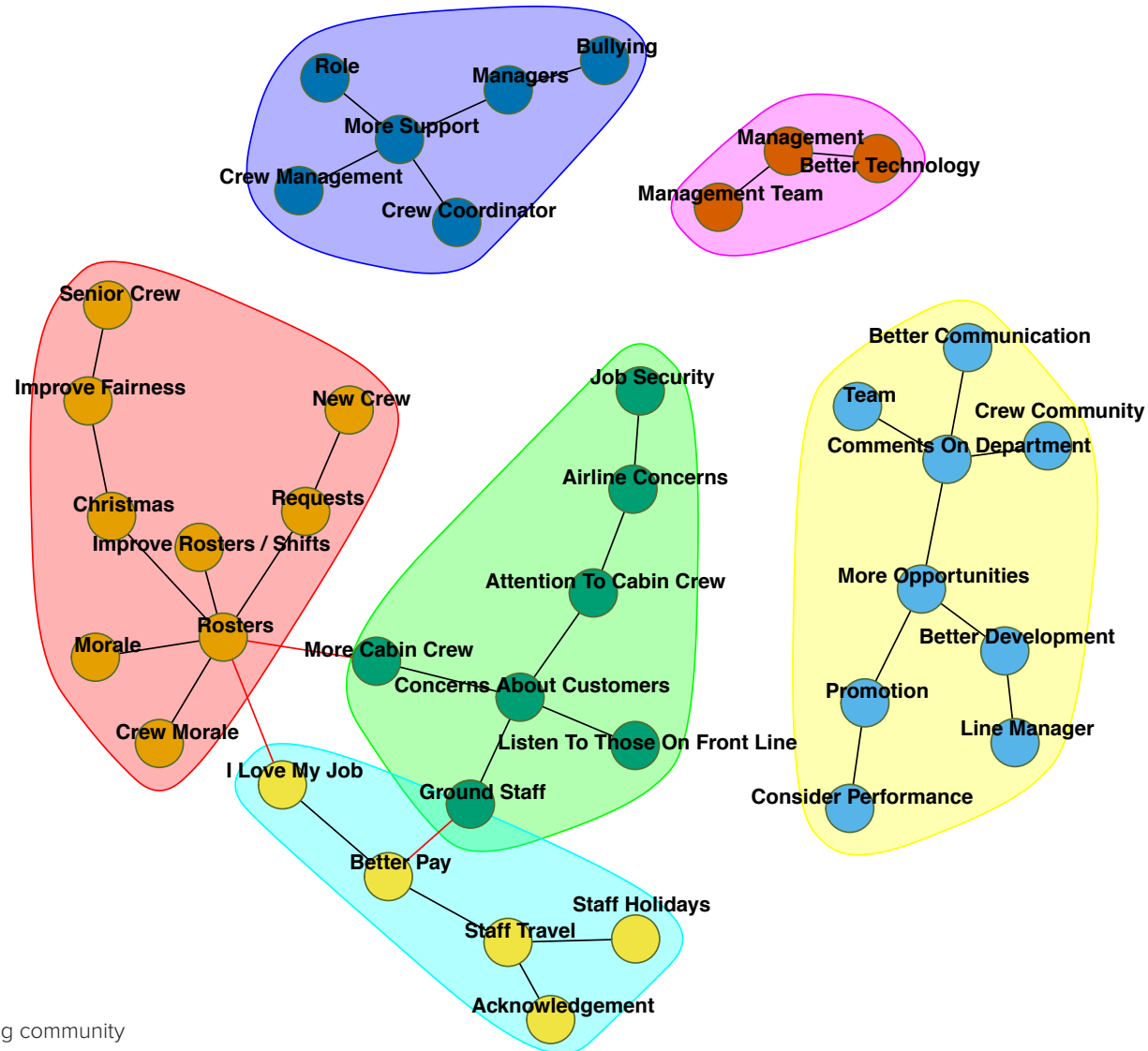
Finding meaningful themes

We aim to highlight meaningful **themes** that increase understanding. We use an **inductive** approach, learning from your data. Our algorithms even learn your business-specific jargon and terminology. The theme labels should be themselves reasonable answers to the question.



Additional insight

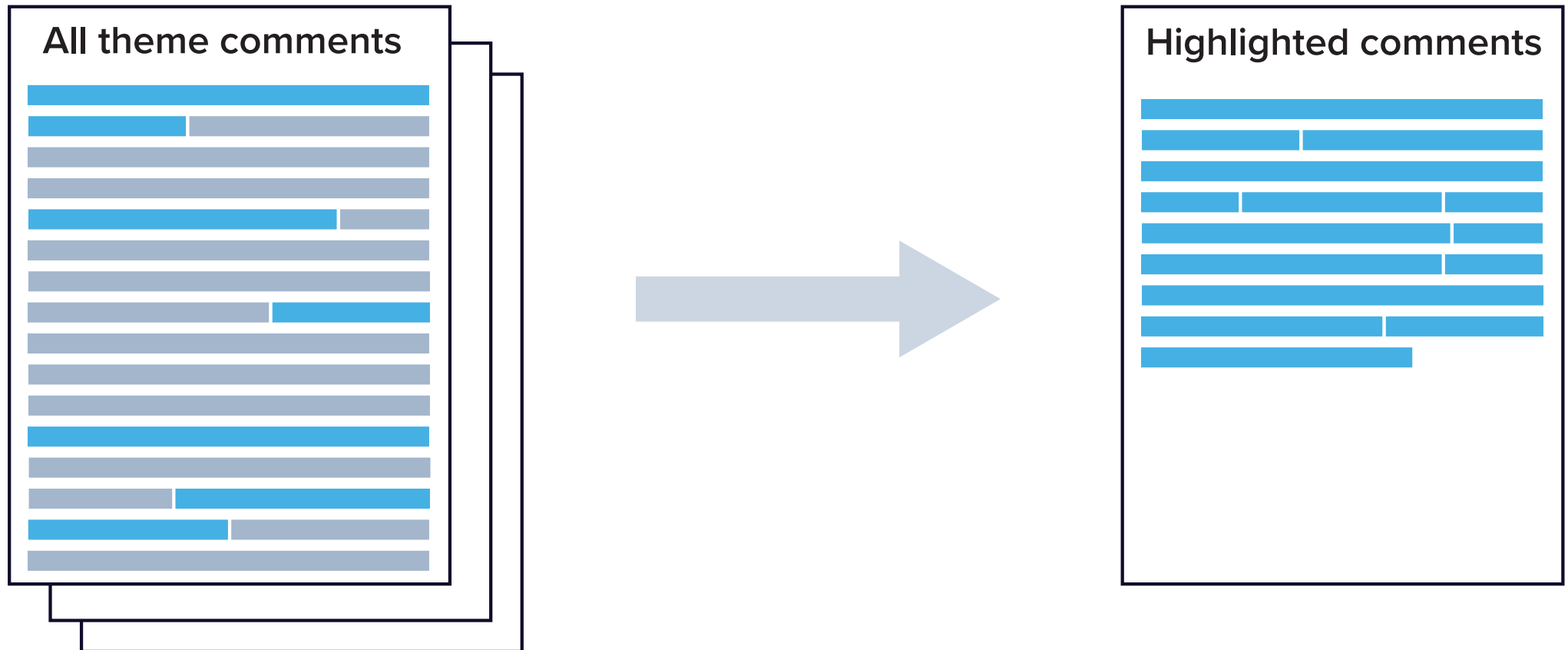
Most clients ask us to identify additional insight from their responses, either looking just at the text data or combining it with other employee information.



A probabilistic co-occurrence network using community detection to identify which themes employees are most likely to talk about together.

Summarising underlying text

We use proprietary document summarisation algorithms to identify example sentences



Most data is unstructured

Most of the talk about Big Data was about the rise in unstructured data

According to Forrester (June 2018) enterprises have huge volumes of
unstructured data

61% > 100 terabytes,

12% > 5 petabytes

Text analysis is everywhere

There are huge numbers of places where text analysis is being used. There are just some of the many common applications

Knowledge
Extraction

Voice interfaces

Chatbots

Spam / Fraud
detection

Search

Advertising,
advertising and
more advertising

Question
answering

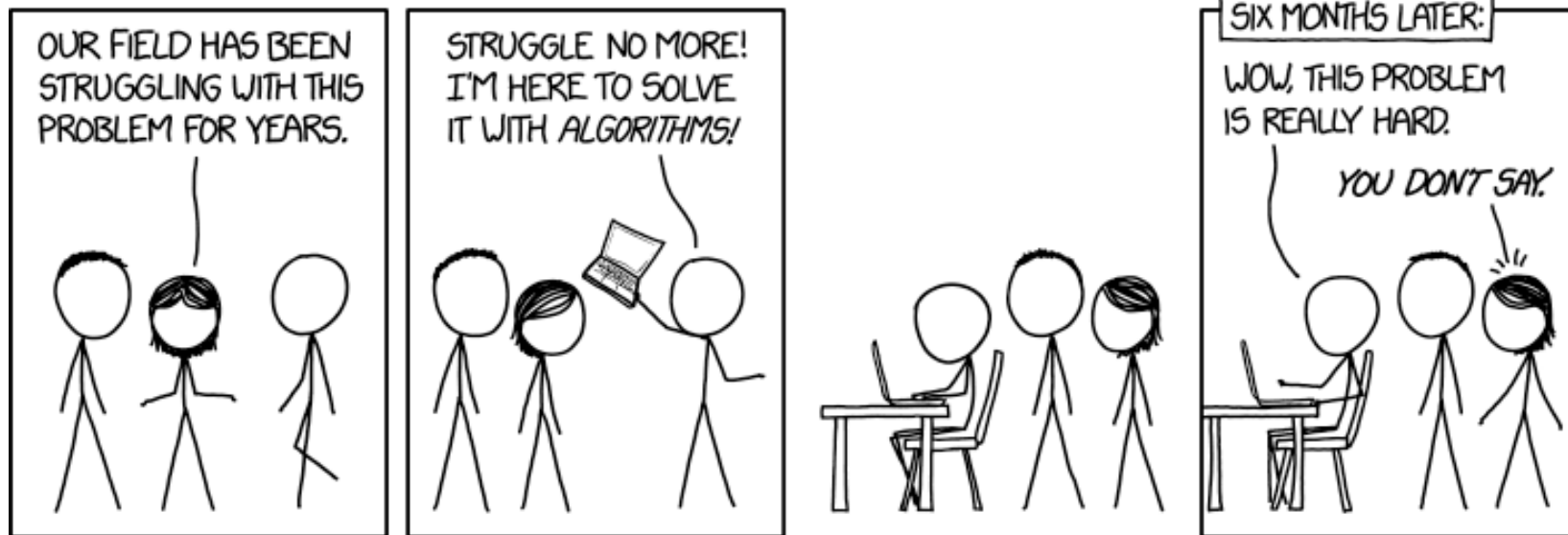
Automated
writing

Spam / Fraud
detection

Patent analysis

There are no easy solutions

But we're getting good at solving specific problems



Translations

[Microsoft 365](#)[Azure](#)[Office 365](#)[Dynamics 365](#)[SQL](#)[Windows 10](#)[More ▾](#)[Research](#)[Research areas ▾](#)[Products & Downloads](#)[Programs & Events ▾](#)[People](#)[Careers](#)[Blogs & Podcasts ▾](#)[Labs & Locations ▾](#)

Achieving Human Parity on Automatic Chinese to English News Translation

March 12, 2018

[Download PDF](#)[View Link](#)[BibTex](#)

Authors

[Hany Hassan Awadalla](#)[Anthony Aue](#)[Chang Chen](#)[Vishal Chowdhary](#)[Abstract](#) [Related Info](#)

Machine translation has made rapid advances in recent years. Millions of people are using it today in online translation systems and mobile applications in order to communicate across language barriers. The question naturally arises whether such systems can approach or achieve parity with human translations. In this paper, we first address the problem of how to define and accurately measure human parity in translation. We then describe Microsoft's machine translation system and measure the quality of its translations on the widely used WMT 2017 news translation task from Chinese to English. We find that our latest neural machine translation system has reached a new state-of-the-art, and that the translation quality is at human parity when compared to professional human translations. We also find that it significantly exceeds the quality of crowd-sourced non-professional translations.

Assistants - Powered by text analysis

The Google Assistant can help you get things done over the phone



0:07 / 1:27



Image to text

Automatically generated text, using a Corpus of Romantic Novels.



Generated story about image

Model: Romantic Novels

"We had to act much of the group, despite the fact that people were surrounded by thousands of people. A few more times, they were beautiful, and I wondered what was going on in New York."

"I didn't want to tear my eyes away from Bill. I had no idea what he meant to me. The rest of the crowd gathered around us, and for a moment. I was pretty sure she was the only person in the entire world. He sounded so relieved and heavy, as if it took hundreds of people standing upon them."

Voice Assistants

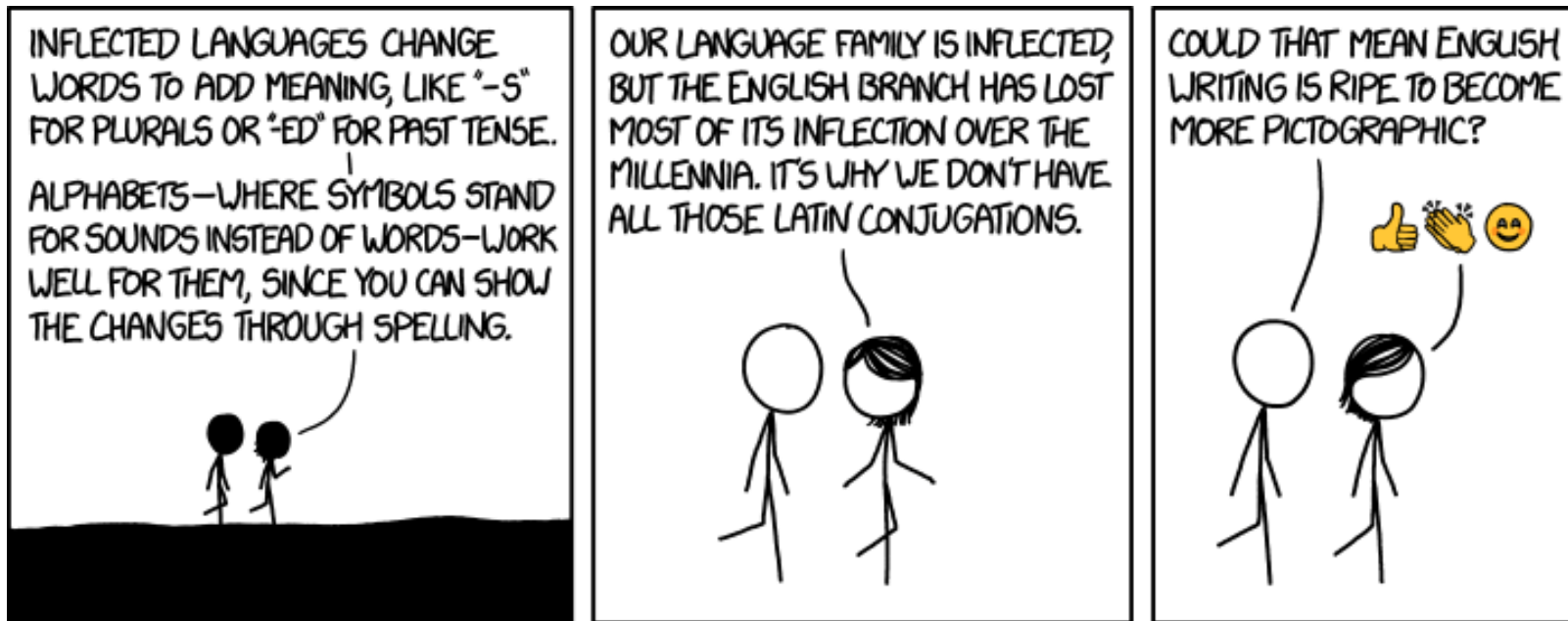
Apple's approach to let Siri do everything is much more difficult to achieve than Amazon's approach to opening Alexa up to developers to build specific models.



“we hve a gud communication with our team member n
also hve a gud cordination with our maneger regarding our
work”

More than just understanding the words

We need to find a way to deal with the way people communicate, especially in short texts.



Need to understand the semantic meaning

“If you could do one thing to improve working for COMPANY what would it be?”

“more money, less hours!”

“Not being salary capped after
30 + years of loyal service”

“Competitive salary increases
for internal movement, low pay
raises encourage outward moves
rather than talent retention”

“Better pay”

“Better wages”

“if I could I would give a pay rise
to the living wage amount and
not just the minimum wage, we
do work hard for our company.”

Need to understand the semantic meaning

“If you could do one thing to improve working for COMPANY what would it be?”

“more money, less hours!”

“Not being salary capped after
30 + years of loyal service”

“Competitive salary increases
for internal movement, low pay
raises encourage outward moves
rather than talent retention”

“Better pay”

“Ha Ha. A raise”

“Better wages”

“if I could I would give a pay rise
to the living wage amount and
not just the minimum wage, we
do work hard for our company.”

Need to understand the semantic meaning

“If you could do one thing to improve working for COMPANY what would it be?”

“Ha ha a raise!!!”

“more money, less hours!”

“Not being salary capped after
30 + years of loyal service”

“Competitive salary increases
for internal movement, low pay
raises encourage outward moves
rather than talent retention”

“Better pay”

“I need a better structure”

“Better wages”

“if I could I would give a pay rise
to the living wage amount and
not just the minimum wage, we
do work hard for our company.”

Parts of Speech tagging (POS)

Understand what each part of a sentence means



There is a reasonably high barrier for analysis to be useful

The more you are going to rely on text, the better the coding has to be. You should be able to understand the comment just by the label.



A word cloud of negative terms for analysis, arranged in a circular pattern. The words are: Useless (blue), Ineffective (green), Wordclouds (blue), Overused (yellow), Rubbish (dark red), trivial (brown), Chartjunk (light green), pointless (green), Superficial (green), and Pretty (orange).

Useless

Ineffective

Wordclouds

Overused

Rubbish

trivial

Chartjunk

pointless

Superficial

Pretty

Very good, at narrow use-cases

The Google Assistant can help you get things done over the phone

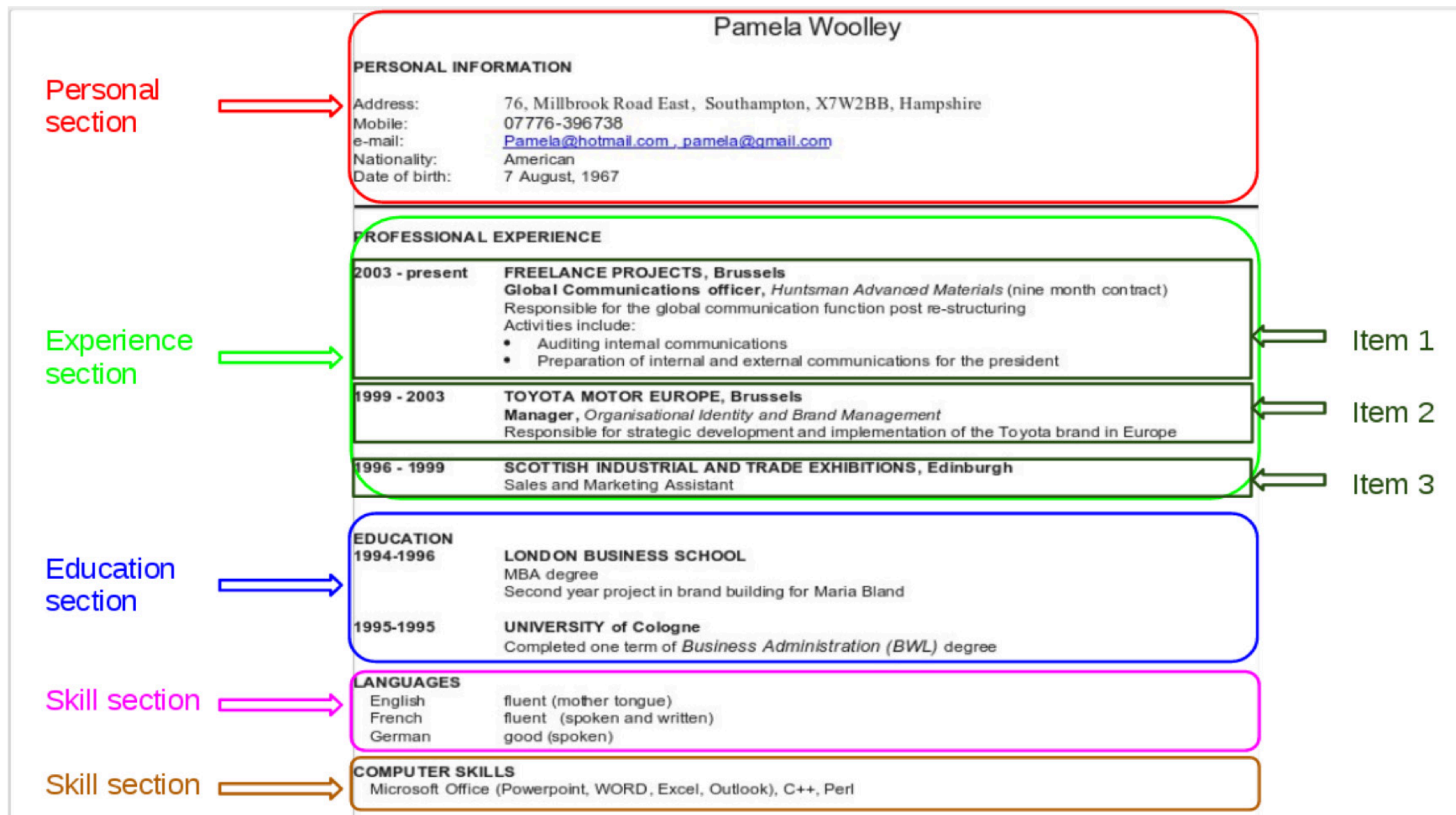


The system only works when it has been trained in extremely narrow contexts, or “domains”, ... since the range of the language used is limited and it is easier to predict what people will say next.

0:07 / 1:27

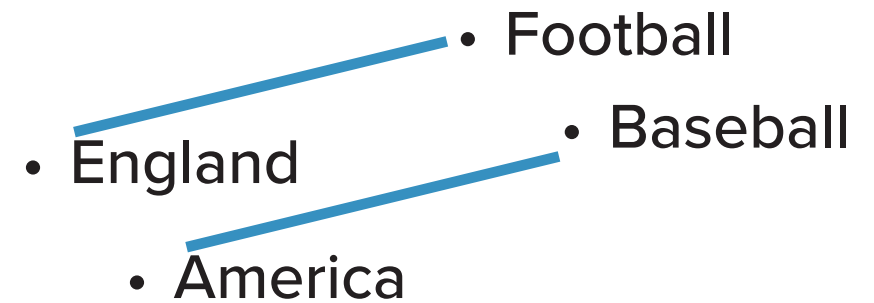


CV / Job Parsing - Importance of 'Zoning'



Text embeddings learns relationships

Vectors can be subtracted to show learned relationships.



England	Football	America	Show	Top N: 2
Baseball America				

BMW	car	Boeing	Show	Top N: 2
airplane aircraft				

Word Embeddings

Dimensionality reduction can give you an idea of what is happening



Text Embeddings

They can do some 'crazy' things

facebook research

Search



Research Areas

Publications

People

Academic Programs

Downloads & Projects

Careers

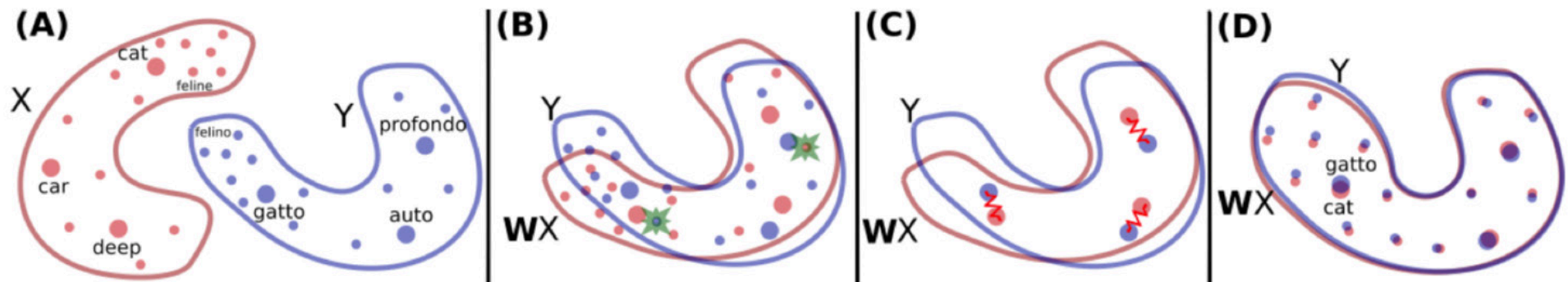
Blog

April 30, 2018

Word Translation Without Parallel Data

International Conference on Learning Representations (ICLR)

By: Alexis Conneau, Guillaume Lample, Marc'Aurelio Ranzato, Ludovic Denoyer, Hervé Jégou



Word embeddings are everywhere

Lots of uses of context words



will liver|

will liver**pool win the league**
will liver**man**
will liver**pool win tonight**
will liver**pool**
will liver**pool win the champions league**
will liver**pool bottle it**
will liver**pool win the league 2019**
will liver**pool win today**
will liver**pool baseball**
will liver **cancer kill you**

Google Search

I'm Feeling Lucky

Report inappropriate predictions
[Learn more](#)

Sentiment Analysis cont.

Current state-of-the-art is about 65% accuracy

Try the API

You always surprise me with your mediocracy

ANALYZE

[See supported languages](#)

Entities

Sentiment

Syntax

Categories

Document & Sentence Level Sentiment

Entire Document

You always surprise me with your mediocracy

Score

Magnitude

0.30.3

0.30.3

Score Range

-1.0 — -0.25

-0.25 — 0.25

0.25 — 1.0

Because of different writing styles, text analytics can suffer bias

For example, regardless of whether it is manager or self-assessment, if a women is involved the text statements are longer.



Ability to make changes quickly

Text, analysed correctly can identify be a real driver for meaningful change. Our approach means we regularly find themes which are unlikely to have been included in a traditional survey.

From...

in the next survey

...To

“Change the music. It’s so moody and depressing and totally not our customer. We have had several complaints from customers and the team”

“Since the last survey, they have changed the music and it creates a better work place for staff and customers”

Andrew Marritt | Founder & CEO

andrew.marritt@organizationview.com | +41 78 905 0775

All statements in this report attributable to Gartner represent OrganizationView GmbH interpretation of data, research opinion or viewpoints published as part of a syndicated subscription service by Gartner, Inc., and have not been reviewed by Gartner. Each Gartner publication speaks as of its original publication date (and not as of the date of this presentation. The opinions expressed in Gartner publications are not representations of fact, and are subject to change without notice.

The Gartner Cool Vendor Logo is a trademark and service mark of Gartner, Inc., and/or its affiliates, and is used herein with permission. All rights reserved. Gartner does not endorse any vendor, product or service depicted in its research publications, and does not advise technology users to select only those vendors with the highest ratings or other designation. Gartner research publications consist of the opinions of Gartner's research organization and should not be construed as statements of fact. Gartner disclaims all warranties, expressed or implied, with respect to this research, including any warranties of merchantability or fitness for a particular purpose.