



machine learning
excellence delivered

Machine Learning (ML) and Artificial Intelligence (AI) for Publishers



Machine Learning, Deep Learning and Data Science become quite popular buzzwords these days. However, question arise: will it change your business?

We try to answer this specifically for the book publishers.

According to some reports, up to 70% of companies will adopt Artificial Intelligence (AI) by 2020.

...Although we find these numbers greatly exaggerated, it is undeniable that AI is gradually changing the way all businesses operate.



Introduction

ML Rule of Thumb



If a typical person can do a mental task with less than one second of thought, we can probably automate it using AI either now or in the near future.

Andrew Ng, Stanford Uni,

What Machine Learning Can Do

A simple way to think about supervised learning.

INPUT A	RESPONSE B	APPLICATION
Picture	Are there human faces? (0 or 1)	Photo tagging
Loan application	Will they repay the loan? (0 or 1)	Loan approvals
Ad plus user information	Will user click on ad? (0 or 1)	Targeted online ads
Audio clip	Transcript of audio clip	Speech recognition
English sentence	French sentence	Language translation
Sensors from hard disk, plane engine, etc.	Is it about to fail?	Preventive maintenance
Car camera and other sensors	Position of other cars	Self-driving cars

- Fraud detection
- Recommendations
- Customer segmentation
- Text analysis
- Text summarization

SOURCE ANDREW NG

© HBR.ORG

Source: [What Artificial Intelligence Can and Can't Do Right Now](#)

Buzzwords



Data science

Big data

Data mining

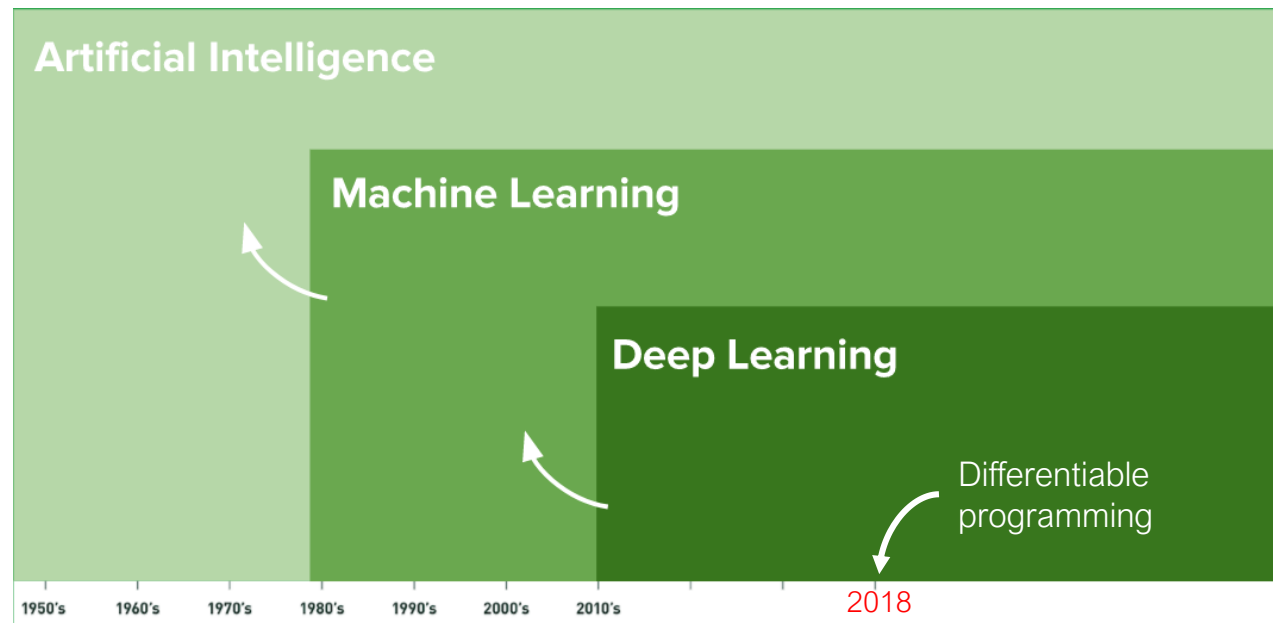
Synthetic intelligence

Cognitive computing

Affective computing

Deep tech

Medtech



Wikipedia:

[Outline of artificial intelligence](#)

Image source: [Jason's Machine Learning 101](#) presentation, Dec 2017

TL;DR



What is / Why use ML?

TL;DR:

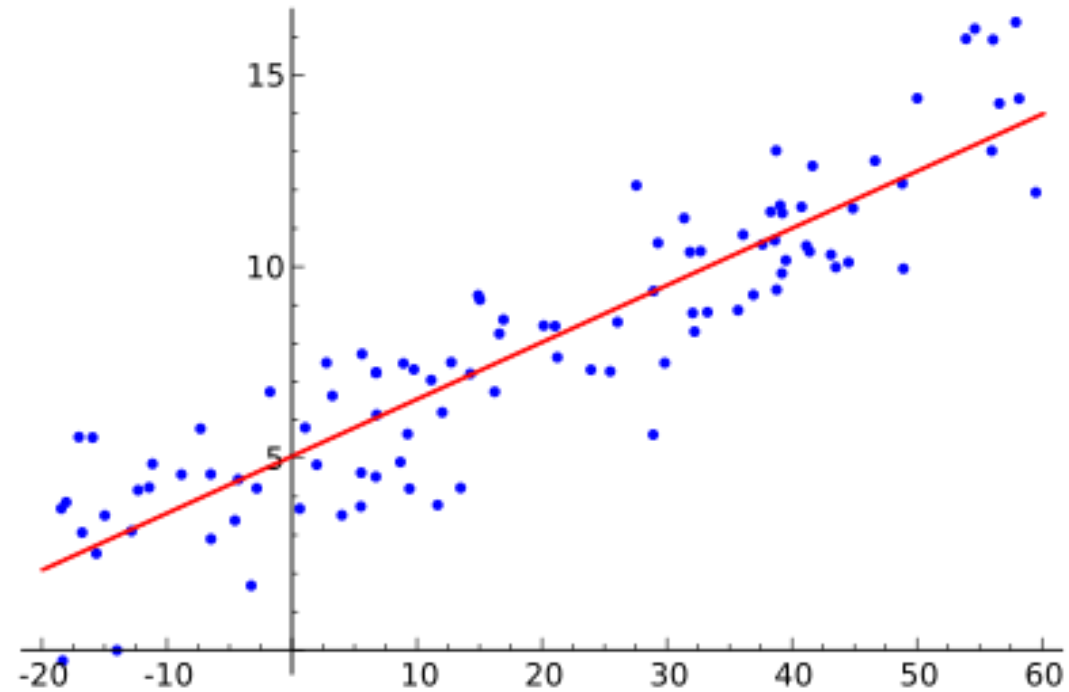
- Software tool for getting answers and decision making
- Mainly based on data, not rules/algorithms
- Two-step:
 - Automatic extraction of “knowledge” (patterns) from past data
 - Prediction on new data

Which data?

- See below...

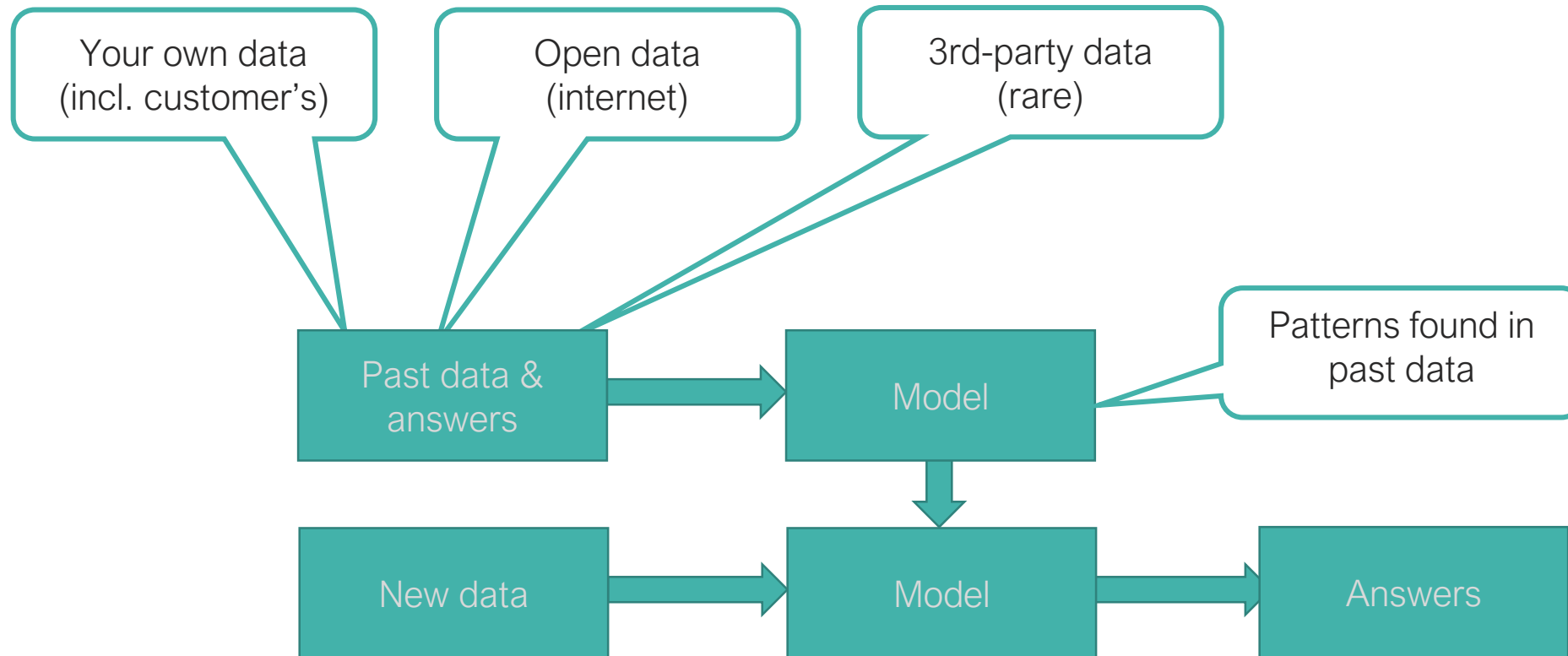
Machine Learning Principle

Easiest ML algorithm

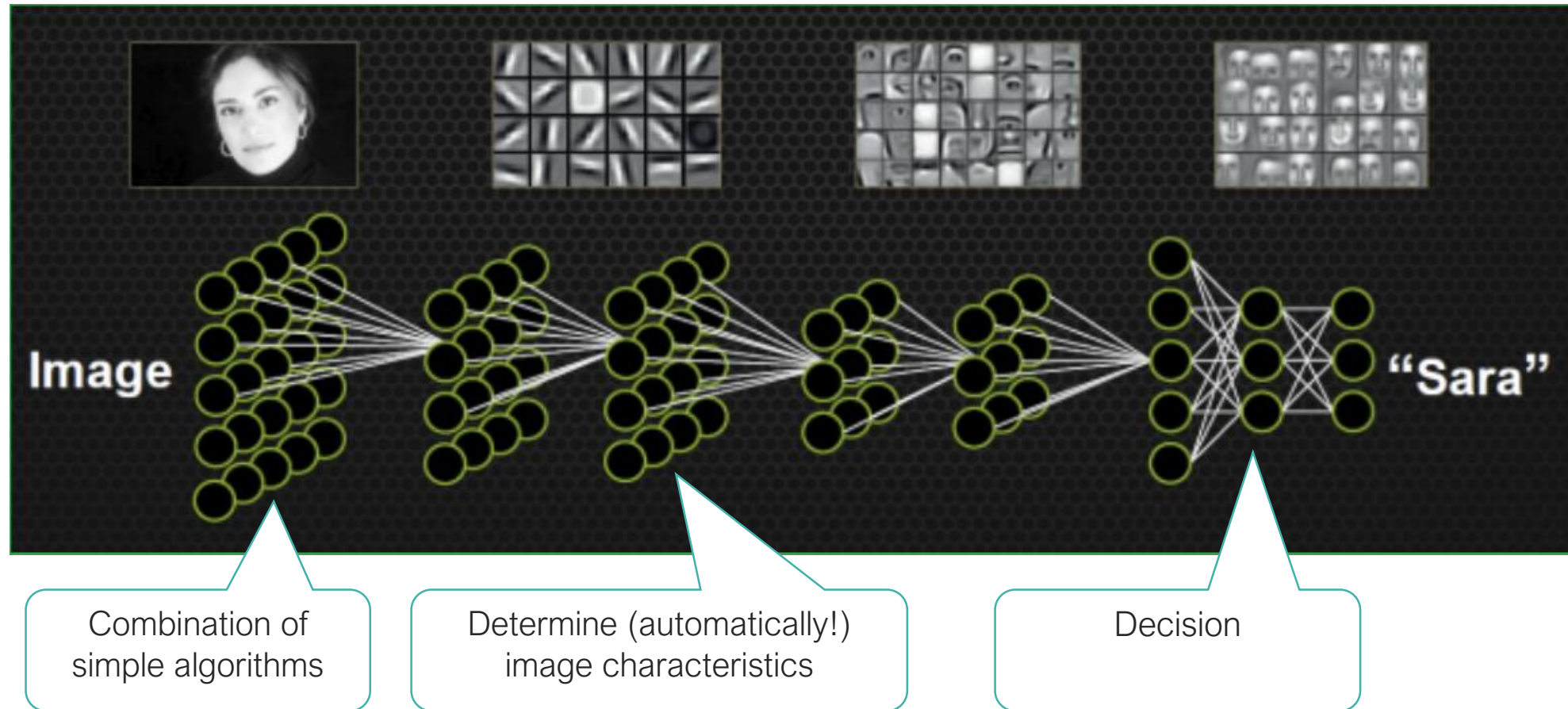


Is my business grow and how?

Data is the king!



Deep Learning: Image Recognition



Deep Learning: revolution

Nearly human performance in

- Image recognition
- Object detection:
 - Images
 - Video
- Translation

2012 contest

ImageNet Challenge

IMAGENET

- 1,000 object classes (categories).
- Images:
 - 1.2 M train
 - 100k test.



DL example: Google Translate ...



“Mongolian”

ЫВО алолд фыалллл оооооооооо
оооооооооооо фЛОЛД игишщгпр
оооооооооооо шгигигшщр оооооооооо



German

?

“Mongolian”

[illegible]

English

?

Google Translate fault (or joke...)



“Mongolian”

ыво алолд фыалллл оооооооооо
оооооооооооо фЛОЛД игишщгпр
оооооооооооо шгигигшщр оооооооооо



German

Du musst dich selbst töten
warf seinen Kopf auf seine Knie

“Mongolian”

[illegible]

English

Have you forgotten that you have not forgiven yourself yet?

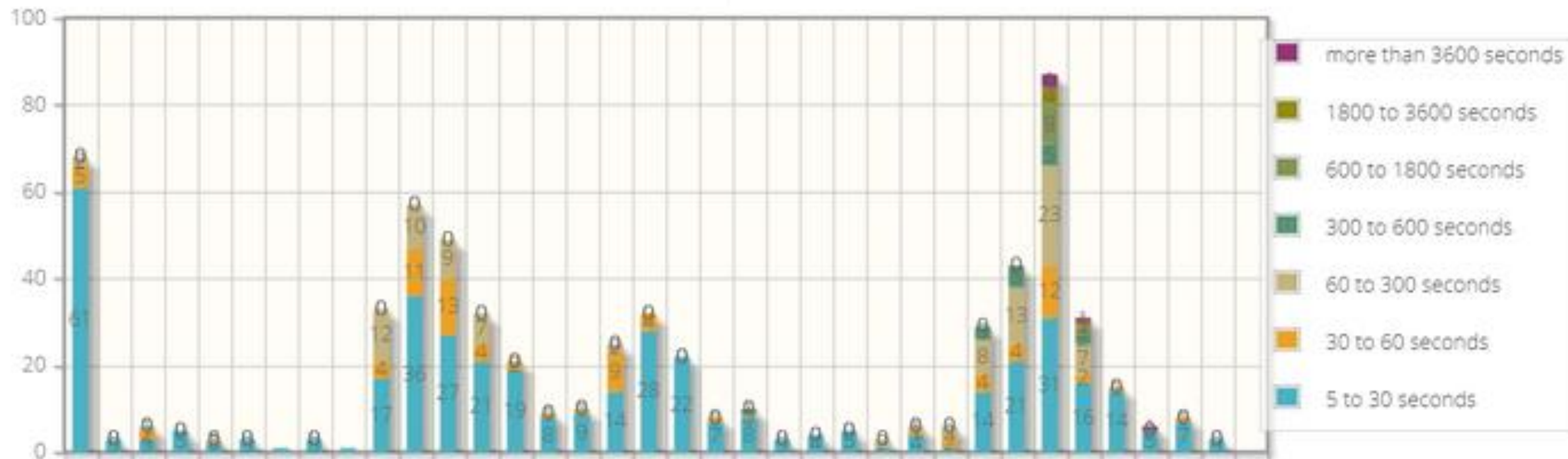
Machine Learning in Publishing

What data to process?



- Structured
 - Sales data
 - Web search statistics, etc.
 - **Access records / reading stats** for eBooks
- Semi-structured
 - Invoices, other business documents
 - Reports, editor reviews
- Free-form
 - Book texts
 - Customer reviews (like Amazon)

Reading statistics (averaged)



Page / Section / Magazine article

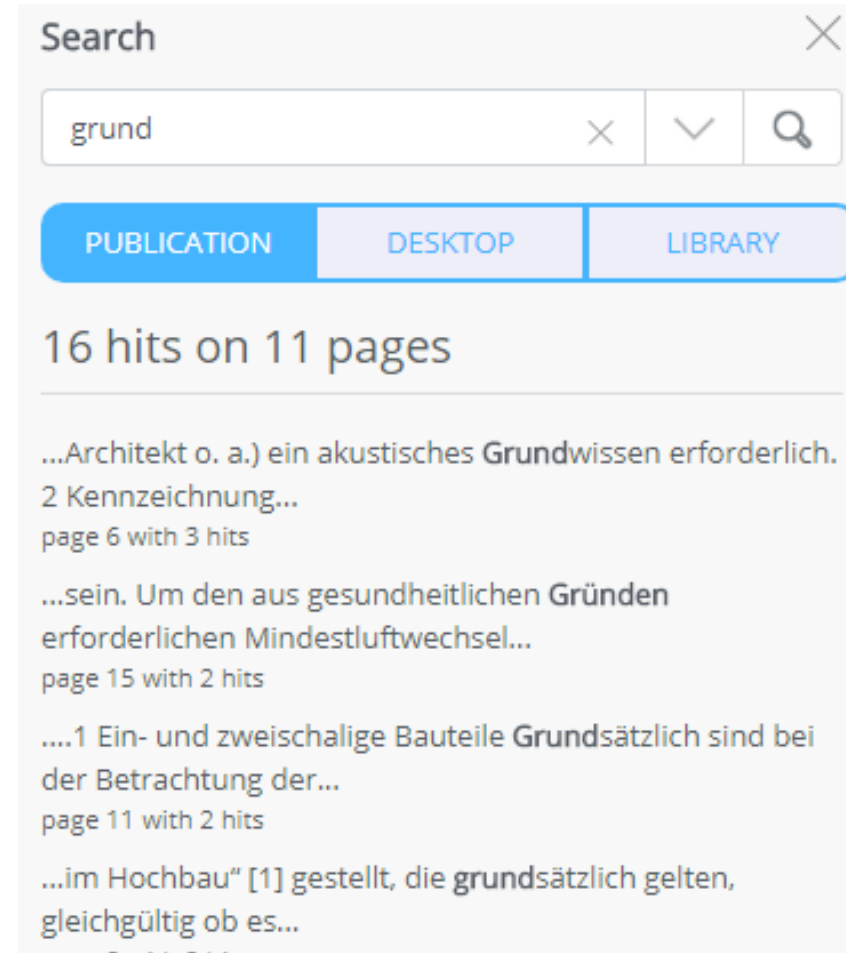
Source: internal SilkCode / LookUP! statistics



Reading statistics: applications

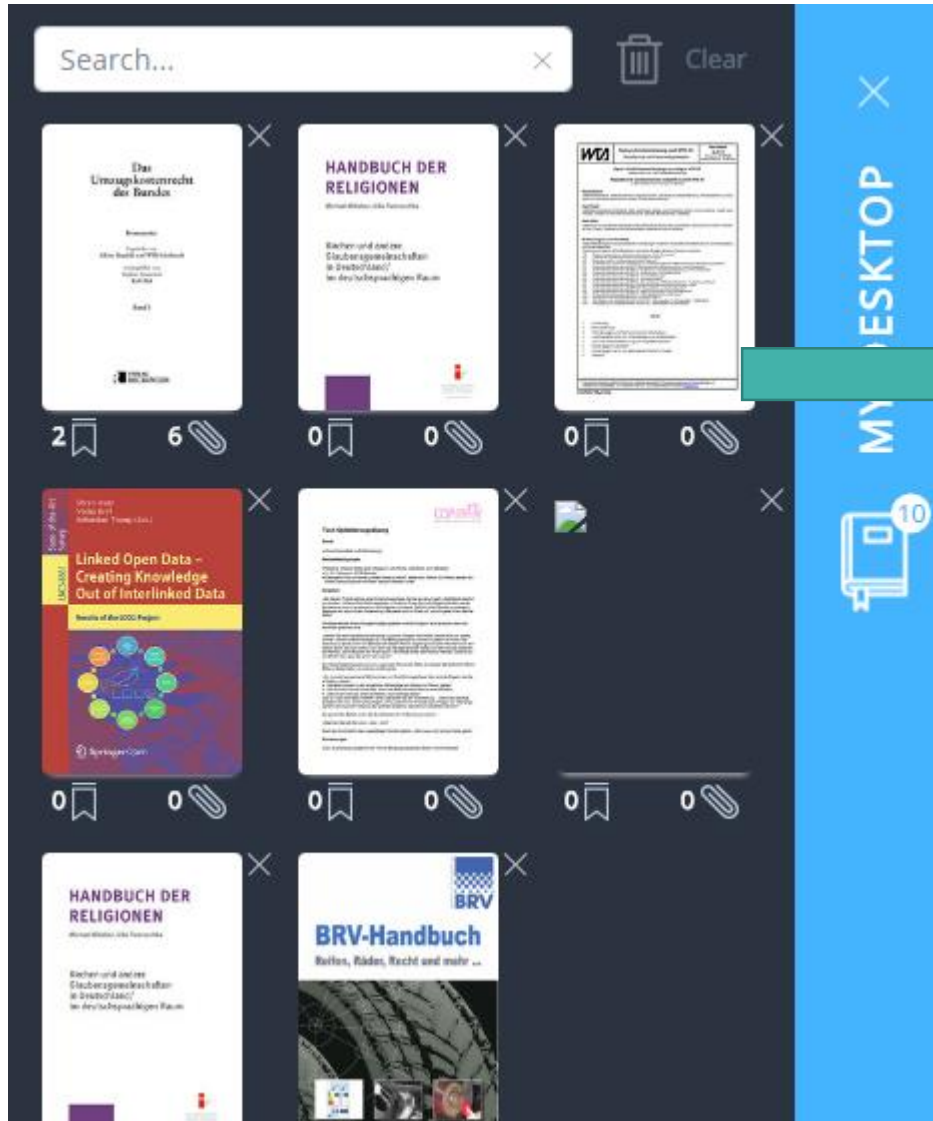


- Recommendations
 - Reading profile analysis
 - Improved by text analysis
 - Recommend similar material
- Search improvements
- Feedback to publisher
- Personalized ads
- Reader's notes / highlights:
 - Interest / importance
 - Complexity (education)

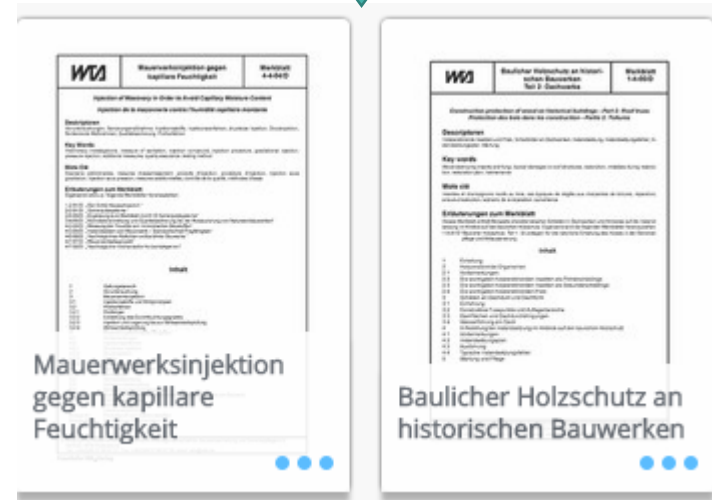


Future: Hits ordering by popularity

Recommendation systems

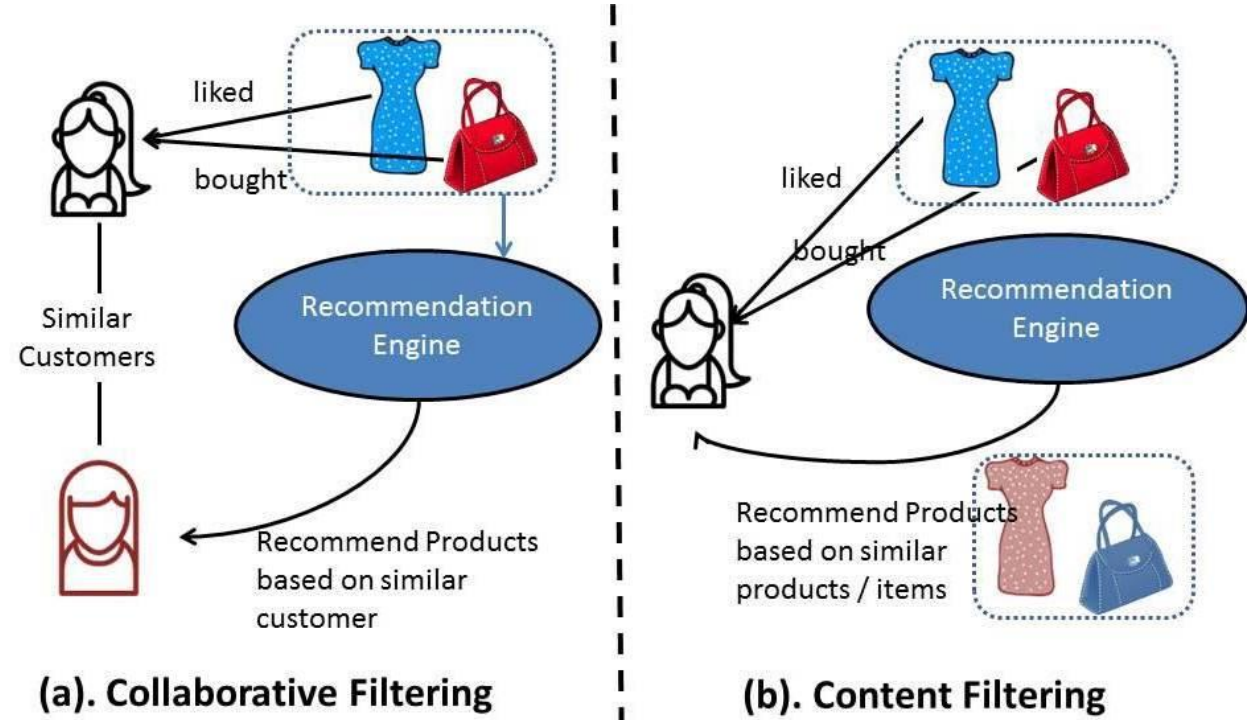


LookUP! perspective:
“You may also be interested in ...”
feature



Recommendation systems

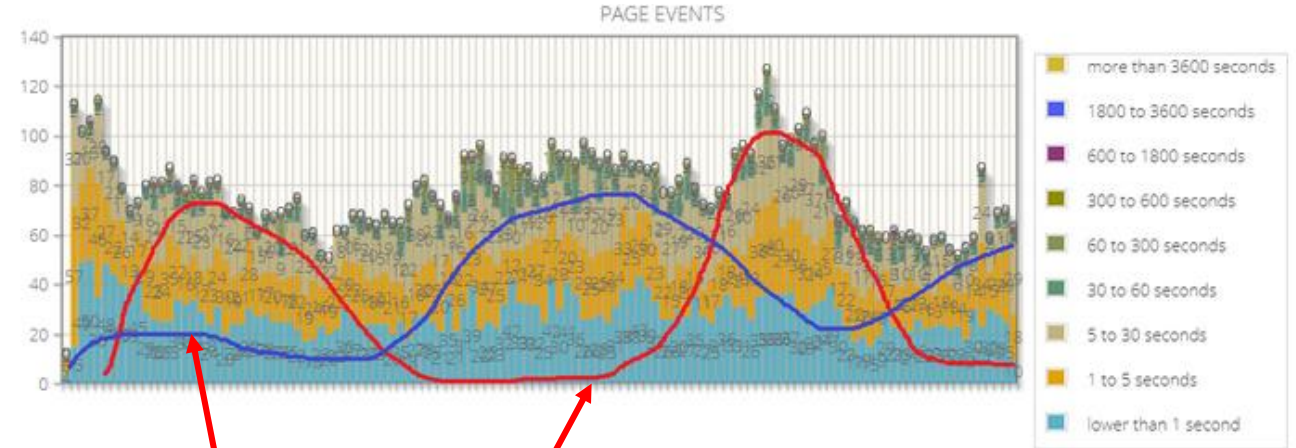
- Recommend book/journal
 - ...similar to a current one
 - ...similar to all material read
 - ...read by users with similar interests
- Analysis of reading habits (profile)



Recommendation systems

Book similarity

- Metadata
 - Author
 - Publisher's descriptions
 - ...
- NLP-extracted keywords
- Applications:
 - Editorial process automation
 - Book recommendation



Readers with different
interests within the same
material

Test data: book plots



Wikipedia ID	1166383
Freebase ID	/m/04cvx9
Book title	White Noise
Book author	Don DeLillo
Publication date	1985-01-21
Genres	Novel, Postmodernism, Speculative fiction, Fiction

Book metadata

Plot summary

Set at a bucolic Midwestern college known only as The-College-on-the-Hill, *White Noise* follows a year in the life of Jack Gladney, a professor who has made his name by pioneering the field of Hitler Studies (though he hasn't taken German language lessons until this year). He has been married five times to four women and has a brood of children and stepchildren (Heinrich, Denise, Steffie, Wilder) with his current wife, Babette.

...

Source: <http://www.cs.cmu.edu/~dbamman/booksummaries.html>

Text similarity explained



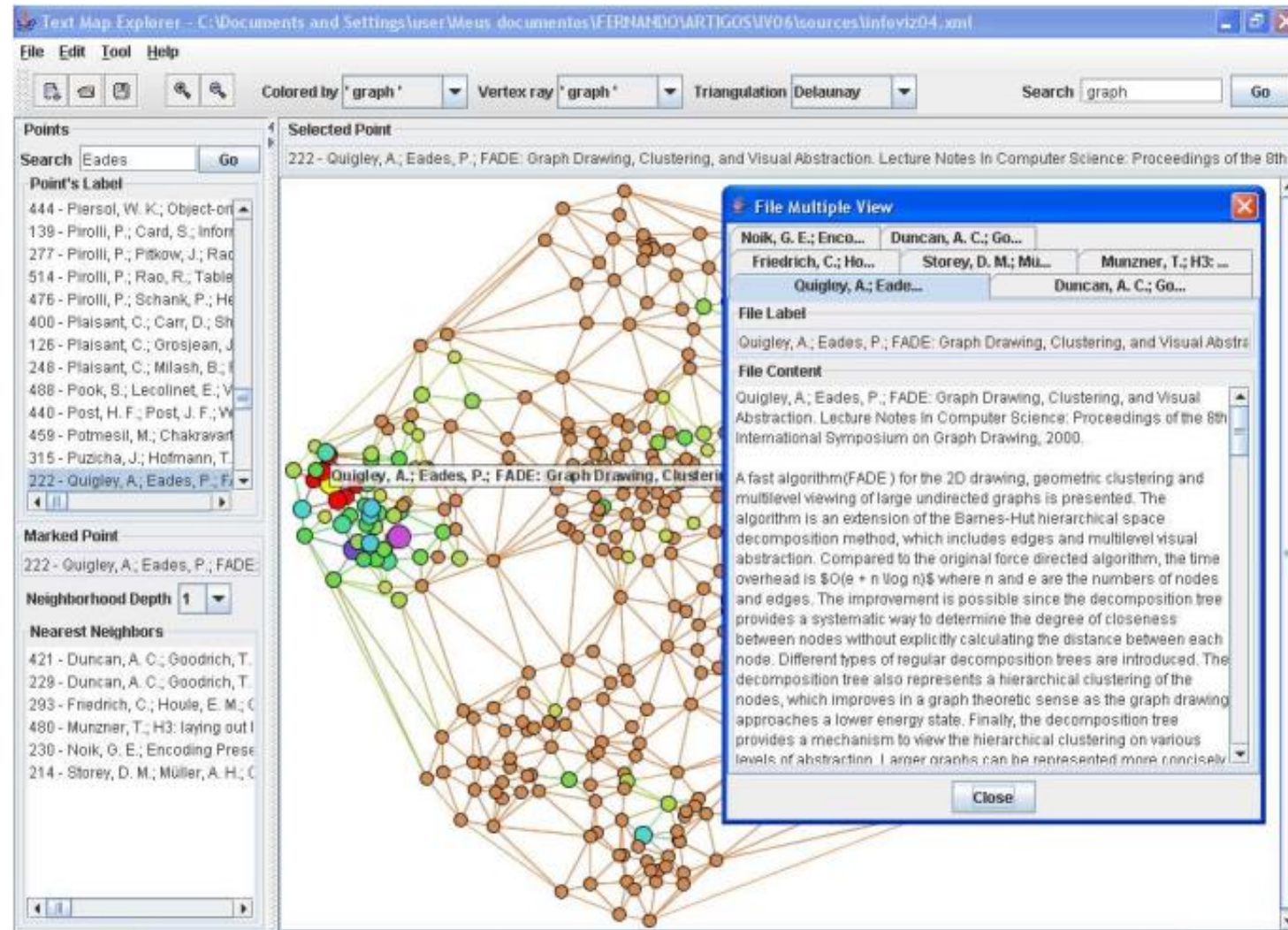
Animal Farm

Old Major the old boar on the Manor Farm calls the animals on the farm for a meeting where he compares the humans to parasites and teaches the animals a revolutionary song Beasts of England When Major dies two young pigs Snowball and Napoleon assume command and turn his dream into a philosophy The animals revolt and drive the drunken and irresponsible Mr Jones from the farm renaming it Animal Farm They adopt Seven Commandments of Animalism the most important of which is All animals are equal Snowball attempts to teach the animals reading and writing food is plentiful and the farm runs smoothly The pigs elevate themselves to positions of leadership and set aside special food items ostensibly for their personal health Napoleon takes the pups from the farm dogs and trains them privately Napoleon and Snowball struggle for leadership When Snowball announces his plans to build a windmill Napoleon has his dogs chase

Snowball's Chance

The story begins with the death of Napoleon the original antagonist of Animal Farm The animals of the farm fearing what will become of them learn that Snowball is alive and well and Snowball returns to the farm to encourage capitalism 0.483 A second windmill is soon built alongside the first and the two are thenceforth known as the Twin Mills allegorical of the Twin Towers of the World Trade Center and the animals all learn to walk on their hind legs something hitherto forbidden by Old Major shortly before the expulsion of Jones from the farm Also in place of the original Seven Commandments Snowball adopts a single slogan for the animals to live by All animals are born equal what they become is their own affair As time passes the animals under the leadership of Snowball realise the properties of monetary gain and begin to file lawsuits against neighbouring farms allowing Animal Farm to gain land and wealth The revitalised farm also attracts animals

Document collection as a map

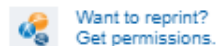


Key sentences

"Our industry is very volatile on the one hand, but on the other hand, if you look into the numbers worldwide, it's one of the most stable industries you can speak of," says 😊 Juergen Boos, president and CEO of the 📖 Frankfurt Book Fair. "And I think Frankfurt reflects this."

RELATED STORIES:

- [PW issue Contents](#)
- More in [International -> Frankfurt Book Fair](#)
- More in articles by [Andrew Albanese](#)



Want to reprint?
Get permissions.

FREE E-NEWSLETTERS

Enter e-mail address

☒ PW Daily ☒ Tip Sheet

[SUBSCRIBE](#) [More Newsletters](#)

Of course, 😊 Boos acknowledges that today such a global approach necessarily "steers toward some political questions"—questions he says 📖 Frankfurt organizers will not shy away from. "I think it's necessary, given the role 📖 Frankfurt plays not only in our industry but in society," he says. "Freedom to publish and 📖 freedom of speech are really the backbone of 📖 Frankfurt. And Frankfurt being the mother of all book fairs, it's the most prominent place to discuss a lot of these issues."

Book publishing is, of course, a mature business in some parts of the world, including 📖 Europe and the 📖 U.S. But globally, 😊 Boos points out, there is strong growth in a number of emerging markets—particularly in Africa, 📖 Asia, and 📖 Latin America. And that growth is fueling an era of international expansion for the 📖 Frankfurt Book Fair.

"At 📖 Frankfurt, over the past 10 years, we went from 60% 📖 German exhibitors, 40% international to the other way around," Boos says. "We now have more international exhibitors than we do 📖 German-language exhibitors—it's like 65% to 35%." And along with that international business comes an equally important opportunity: to showcase more talent, stories, and cultures from around the world—something fair organizers are always working to encourage. "It is very important to listen to other voices," Boos adds, "and I think we must invest as an industry to make these voices accessible."

Frankfurt Book Fair 2018 Preview

Spanning the globe, the Frankfurt Book Fair shows that publishing is still a growth business

By Andrew Albanese | Sep 21, 2018



It's a common question in our constantly evolving digital world: are publishers adapting quickly enough? Yet through a decade of rapid technological change, a global recession, and ever-shifting consumer expectations, the global book business has shown its resilience. And one reason has to do with the 📖 Frankfurt Book Fair, an example of the industry's strength.

"Our industry is very volatile on the one hand, but on the other hand, if you look into the numbers worldwide, it's one of the most stable industries you can speak of," says 😊 Juergen Boos, president and CEO of the 📖 Frankfurt Book Fair. And I think Frankfurt reflects this.

RELATED STORIES:

[PW issue Contents](#)
More in [International -> Frankfurt Book Fair](#)
More in articles by [Andrew Albanese](#)

Want to reprint?
Get permissions.

FREE E-NEWSLETTERS

Enter e-mail address

☒ PW Daily ☒ Tip Sheet

[SUBSCRIBE](#) [More Newsletters](#)

Of course, 📖 Boos acknowledges that today such a global approach necessarily "steers toward some political questions"—questions he says 📖 Frankfurt organizers will not shy away from. "I think it's necessary, given the role 📖 Frankfurt plays not only in our industry but in society," he says. "Freedom to publish and 📖 freedom of speech are really the backbone of 📖 Frankfurt. And Frankfurt being the mother of all book fairs, it's the most prominent place to discuss a lot of these issues."

Indeed, last year, with nationalist movements on the rise in 📖 Europe, 📖 French president Emmanuel Macron joined 📖 German president Angela Merkel at the fair, and the two leaders spoke of a 📖 Europe united—rather than divided—by culture. 📖 Macron told reporters that books and literature are vital—even when the ideas expressed make us uncomfortable. "This is what builds us together—that prevents us from closing ourselves off, from giving way to fear, tribalism, and chaos," he said. To 📖 Merkel's remarks, she added that "freedom of expression gives birth to growth with political freedom."

Politics will likely loom large again at the 📖 Frankfurt Book Fair in 2018, including for 📖 U.S. publishers. On one hand, they're concerned about Trump's trade policies and his attacks on the press. On the other hand, sales of political books are soaring. Last week, 📖 NPR's 📖 Kristin McLean reported that "blockbuster" nonfiction titles have led to growth in print-unit sales through the first half of 2018, with political books posting their highest sales volume in 15 years. 📖 In fact, fiction print-unit sales, however, are down 4%, and over the past three years have declined roughly 11%.

This year's fair again features a strong program, and much of it has political overtones. Among the highlights, award-winning 📖 Nigerian author Chimamanda Ngozi Adichie (📖 American) will give the keynote address on October 9 at the fair's opening press conference, which will be set in the new 📖 Frankfurt Pavilion, situated at the center of the fair's square. Later that day, Adichie will be in 📖 London, where she will be honored with the 📖 CBE Peter Iyia, given annually to a writer who strives to "define the real truth of our times and our societies." 📖 On the fair's opening day, October 10, the annual CEO talk will feature 📖 Macmillan CEO John Sargent, who in January made headlines for defying Trump's effort to stop publication of Michael Wolff's 📖 *Fire and Fury*.

Of course, the backbone of the 📖 Frankfurt Book Fair is the rights trade. And this year, the 📖 U.S. rights agents and editors conduct their business, will once again grow, featuring 520 tables, up from 500 last year, with 350 agencies from around the world. "The traditional rights business is growing, and on top of that we are seeing a lot of licensing and merchandising activities," Boos says. "We believe this is not going to stop."

For details on specific programs, check out the 📖 Frankfurt Book Fair website. You can follow all the action from 📖 Frankfurt through 📖 PW's 📖 Frankfurt Show Daily, available in print at the show, as well as on the 📖 PW website.

The Big Will U.S. Agencies Will Be Selling at the 2018 Frankfurt Book Fair



Named entities

"Our industry is very volatile on the one hand, but on the other hand, if you look into the numbers worldwide, it's one of the most stable industries you can speak of," says 😊 Juergen Boos, president and CEO of the 💎 Frankfurt Book Fair. "And I think Frankfurt reflects this."

Persons

RELATED STORIES:

- [PW issue Contents](#)
- More in [International](#) -> [Frankfurt Book Fair](#)
- More in articles by [Andrew](#)

Book publishing is, of course, a mature business in some parts of the world, including 📌 Europe and the 📌 U.S. But globally, 😊 Boos points out, there is strong growth in a number of emerging markets—particularly in Africa, 📌 Asia, and 📌 Latin America. And that growth is fueling an era of international expansion for the 💎 Frankfurt Book Fair.

Places

Events etc.

Named Entities

MARCUS® 2016, 23/2...

FRIEDRICH-ADOLF-RICHTER-SCHULE
Freie Gemeinschaftsschule der AWO Rudolstadt

DAS LEHRERINTERVIEW

FRIEDRICH-ADOLF-RICHTER-SCHULE
Klassenlehrer der Klasse 11
Jenny B. (Schülerin) 2016

Wo ist Ihr Lieblingsort in der Schule?
Der Klassenraum meiner Klasse, der von den Schülern und mir so gestaltet wurde, dass er für eine angenehme Lernatmosphäre sorgt. Eben einfach ein Ort zum Wohlfühlen.

Was liegt Ihnen an der Schule besonders am Herzen?
Ein lehrerfündliches Klima, eine ganzheitliche Schulleitung, gegenseitige Wertschätzung, Vertrauen, Empathie sowie Ehrlichkeit.

Was würden Sie gerne lernen?
Stille.

Welches Buch lesen Sie gerade?
„Ein ganzes halbes Jahr“ von Jop Heyen.

Was tun Sie in Ihrer Freizeit?
Familie und Freunde treffen, Kochen sowie Reisen.

Was sollte jeder Mensch lernen?
„Nirgends verdrängen will, wird auch das verlieren, was er bezaubern möchte“ (Klausur Hensmann)

Soziale Dienste Rudolstadt gGmbH
Freie Gemeinschaftsschule

Neue Schulstraße 21
07407 Rudolstadt
Telefon: 03672-82 90 180
Fax: 03672-82 90 181
E-Mail: info@friedrich-adolf-richter-schule.de

187 SCHÜLER | 26 LEHRKRÄFTE UND MITARBEITER | 24.000 m² GROSSES SCHULGELÄNDE | 3 LERNHÄUSER JEWEILS FÜR DIE KLASSEN 5 + 6, 7 + 8, 9 + 10 + 11 | THÜRINGEN | AULAMENSA

Anmeldungen für das Schuljahr 2017/18 sind noch möglich.

Tag der offenen Tür:
Fr 09.12.2016 14 - 18 Uhr
Lebendiger Adventskalender:
Fr 09.12.2016 18 Uhr

Schnuppertag für Viertklässler:
Do 12.01.2017 07.00-15.30 Uhr
Führungen u. Informationen zum Schulkonzept:
Sa 14.01.2017 11 - 14 Uhr
Sa 18.02.2017 11 - 14 Uhr

Rudolstadt ist eine Stadt im Landkreis Saalfeld-Rudolstadt im Freistaat Thüringen, Deutschland. Die ehemalige fürstliche Residenz Rudolstadt liegt eingebettet in einem waldumgebenen Tal und zieht sich bandartig am weiten Bogen des Flusses Saale entlang.



Wikipedia

LookUP! perspective:
Automatic Wikipedia
Linking

Sentiment analysis



Quickly find what readers think about your books

Analyze Sentiment

Language
english ▼

Enter text

I used to buy a copy of The New York Times Book Review every week to keep up with the new books being published. When I moved to a small town I lost touch with the publication, and I missed it. With the advent of the internet, I thankfully renewed contact, but I missed being able to settle down in a comfortable chair and getting acquainted with many of the new fiction and non-fiction works.

Enter up to 50000 characters

Analyze

Sentiment Analysis Results

The text is **pos**.

The final sentiment is determined by looking at the classification probabilities below.

Subjectivity

- neutral: 0.2
- polar: 0.8

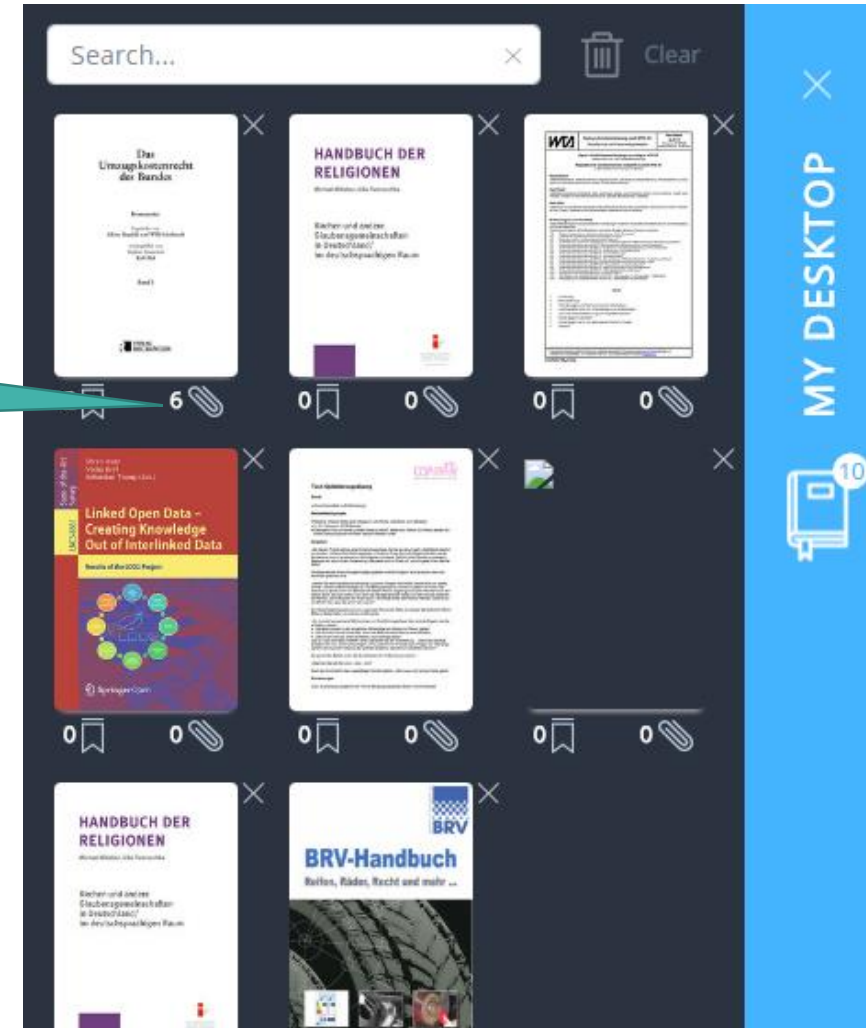
Polarity

- pos: 0.7
- neg: 0.3

Sentiment analysis

LookUP! perspective:
Notes analysis

- ✗ ... seltsam ...
- ✓ Wichtig ...
- ✗ Verstehe nicht
- ✓ Sic!





www.silkdata.ai
hello@silkdata.ai
+49 (2151) 387 3531

Headquarters:
Silk AI lines LLC

Hikaly 3-7
220005 Minsk
Belarus

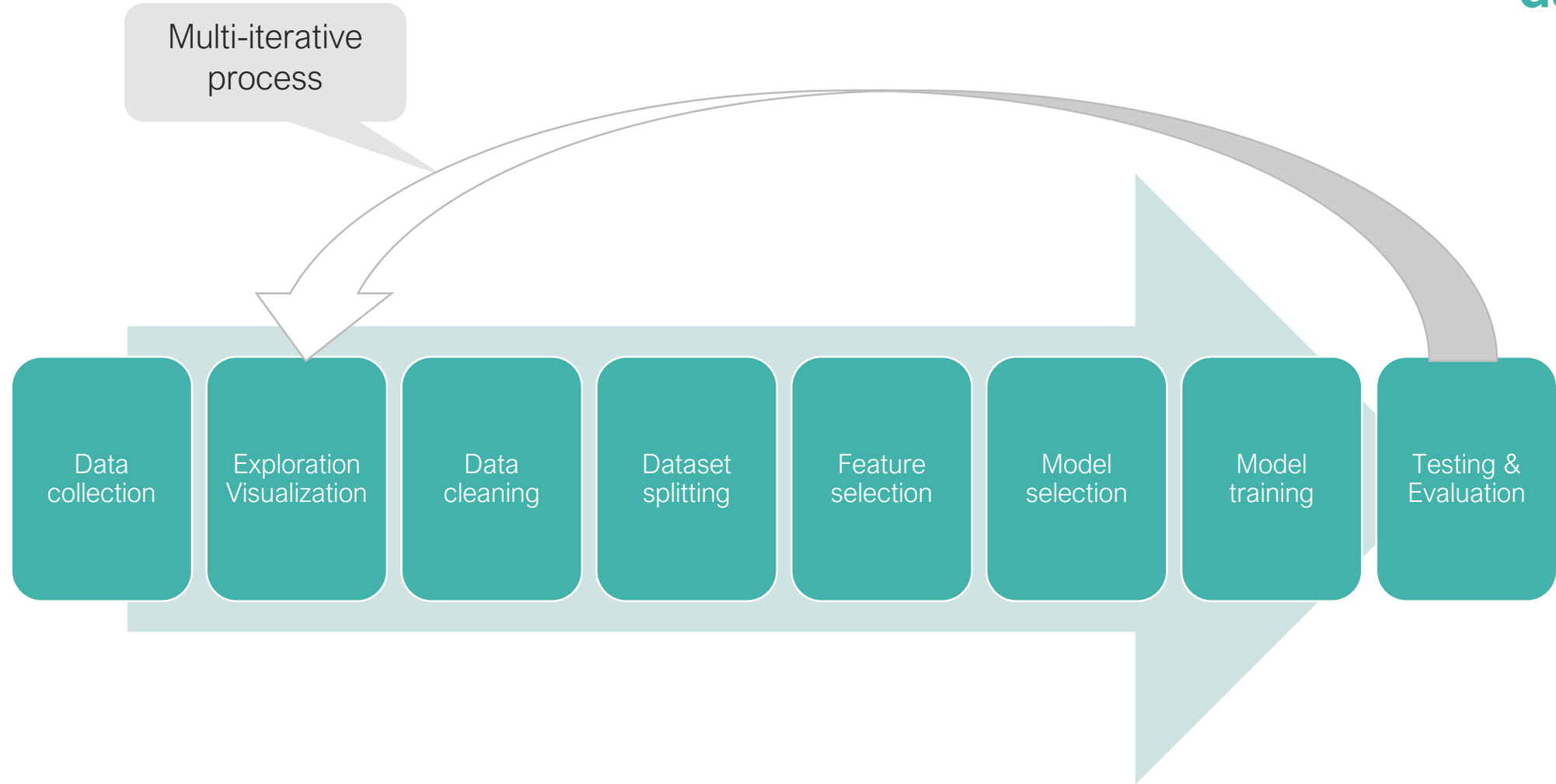
Partner in EU:
SilkCode GmbH

Luisenstraße 62
D-47799 Krefeld
Germany



EXTRA SLIDES

More details of ML pipeline



Text processing pipeline



- Input format analysis and conversion (HTML, PDF, ...)
- Language identification
- Sentence segmentation
- POS-tagging and lemmatization
- Stop-word removal
- ML models
 - TF-IDF
 - LDA and friends
 - Word embedding
 - NER
 - Summarization
 - ...
- Display of result within original text

Approaches and business values



- Structured data
 - Prediction (market, customer churn)
 - Book recommendation
- Semi-structured
 - Document processing automation
 - Document management
 - Data extraction
- Unstructured data
 - NLP (natural language processing)
 - Information extraction
 - Content similarity

Topic analysis: IT Vacancy ads

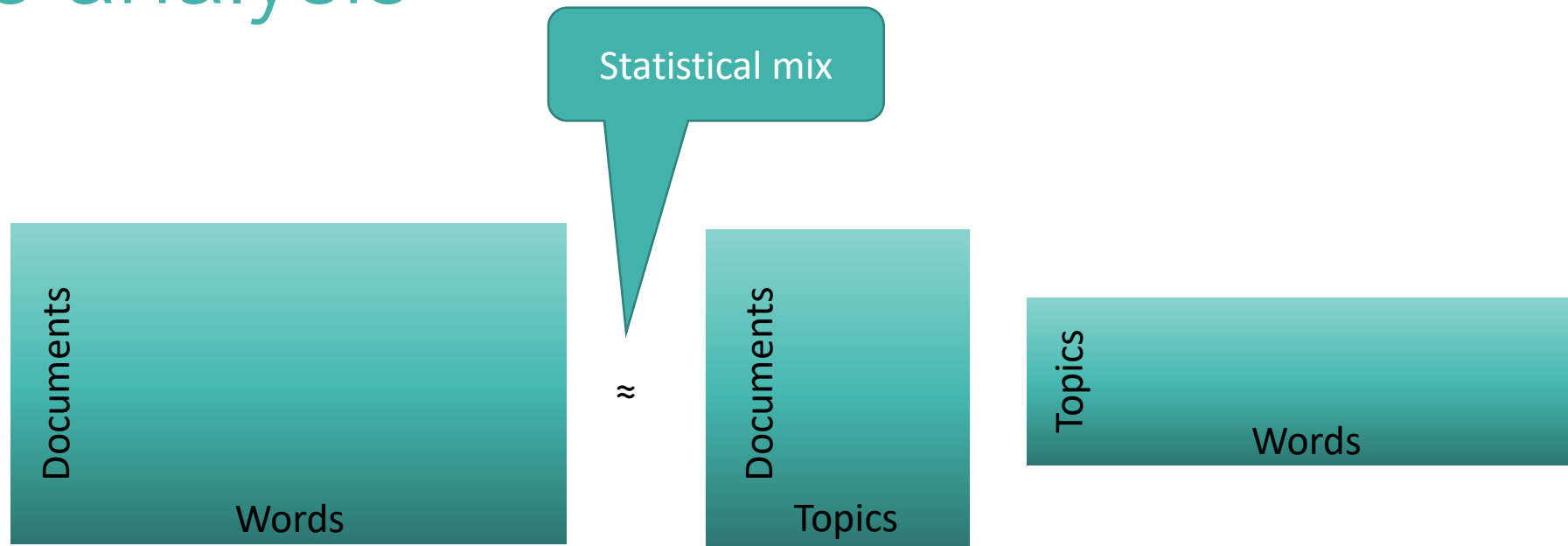


Preprocessing issues

Special terms

- .net, asp.net
 - angular.js
 - *nix (= Unix and/or Linux)
 - C++ / C#
 - SAP R/3, B2B, 3D, 7z
- All kinds of noise
 - Quotes
 - Ligatures
 - Dashes
 - Complex Unicode
 - Always mix with English

Topic analysis



Technical terms

Process

[50.2%] php, mysql, html_cs, lamp, php_mysql, mvc, framework, ...
[13.1%] code, tdd, framework, unit, continuous_integration, apis, practice
[9.7%] tester, startup, manual, like
[8.2%] marketing, seo, campaign, search, strategy, digital_marketing, ...

Marketing

Topic modeling example



Exploratory navigation over Wikipedia

The Hitchhiker's Guide to the Galaxy

related topics

- {film, series, show}
- {work, book, publish}
- {album, band, music}
- {system, computer, user}
- {math, energy, light}
- {theory, work, human}
- {@card@, make, design}

The Hitchhiker's Guide to the Galaxy is a science fiction comedy series created by English writer, dramatist, and musician Douglas Adams. Originally a radio comedy broadcast on BBC Radio 4 in 1978, it was later adapted to other formats, and over several years it gradually became an international multi-media phenomenon. Adaptations have included stage shows, a "trilogy" of five books published between 1979 and 1992, a sixth novel penned by Eoin Colfer in 2009, a 1981 TV series, a 1984 computer game, and three series of three-part comic book adaptations of the first three novels published by DC Comics between 1993 and 1996. There were also two series of towels, produced by Beer-Davies, that are considered by some fans to be an "official version" of *The Hitchhiker's Guide to the Galaxy*, as they include text from the first novel.^{[1][2]} A Hollywood-funded film version, produced and filmed in the UK, was released in April 2005, and radio adaptations of the third, fourth, and fifth novels were broadcast from 2004 to 2005. Many of these adaptations, including the novels, the TV series, the computer game, and the earliest drafts of the Hollywood film's screenplay, were done by Adams himself, and some of the stage shows introduced new material written by Adams.

The title is the name of a fictional, eccentric, electronic travel guide, *The Hitchhiker's Guide to the Galaxy*, prominently featured in the series.

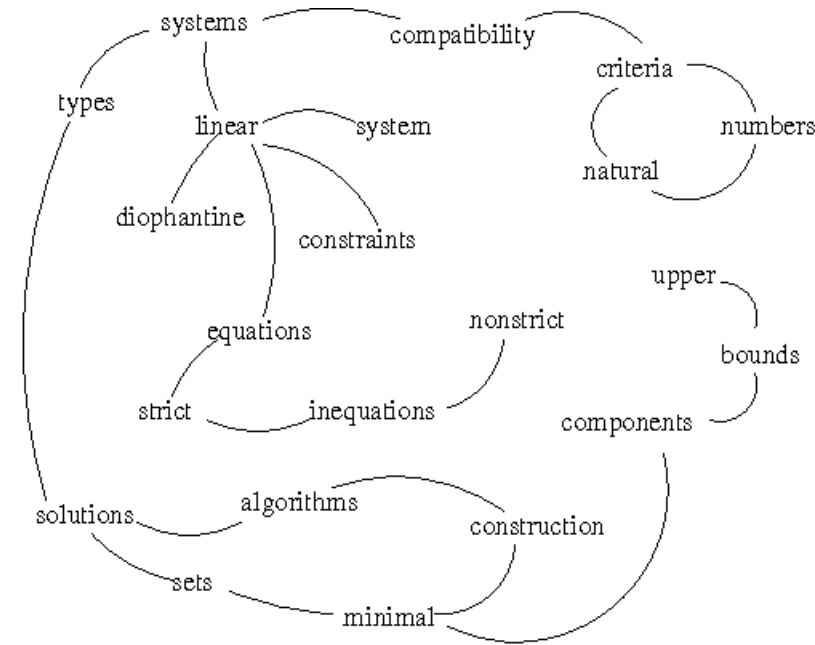
The Hitchhiker's Guide to the Galaxy^[3] is often abbreviated "HHGTTG" (as

related documents

- The Prisoner
- Jack Benny
- Sunset Boulevard (film)
- Superhero
- Buffy the Vampire Slayer (TV series)
- Alfred Hitchcock
- EastEnders
- Cinema of the United Kingdom
- Twin Peaks
- Psycho (1960 film)
- The Kids in the Hall
- Dubbing (filmmaking)
- Humphrey Bogart
- Scooby-Doo
- You Can't Do That on Television

Source: http://www.princeton.edu/~achaney/tmve/wiki100k/docs/The_Hitchhiker's_Guide_to_the_Galaxy.html

- Known usage: O'Reilly



linear constraints; linear diophantine equations; natural numbers; nonstrict inequations; strict inequations; upper bounds

linear constraints; linear diophantine equations; minimal generating sets; non-strict inequations; set of natural numbers; strict inequations; upper bounds

Data Obfuscation: Example



Allstate's contest on Kaggle:

- <https://www.kaggle.com/c/allstate-claims-severity/data>
- “Each row in this dataset represents an insurance claim. You must predict the value for the 'loss' column. Variables prefaced with 'cat' are categorical, while those prefaced with 'cont' are continuous.”

Obfuscation

- No real names / categorical values
- Possible linear mix / rescaling of numerical values
- **Patterns** still remain and can be extracted by model!!!

id	cat1	cat2	cat3	cat4	cat5	cat6	cat7	...	cat116
1	A	B	A	A	A	A	B	...	LB
cont1	cont2	cont3	cont4	cont5	cont6	cont7	cont8	...	loss
0.7263	0.2459	0.1875	0.7896	0.31006	0.71836	0.3350	0.3026	...	2213.18

Platforms and libraries



- Python
 - Gensim
 - SpaCy
 - NLTK
 - patterns
- Java
 - OpenNLP
- Deep Learning
 - Caffe
 - Keras / Tensorflow
- Client solutions: browser or mobile