



THE OPENMINTED PROJECT

OPEN INFRASTRUCTURE FOR MINING TEXT
AND DATA

THE PROBLEM

MOUNTAINS OF DATA ARE BEING PRODUCED

2X

The digital universe – or the data we create and copy annually – is **doubling** in size every **2 years**.¹

44
zetta
bytes

By **2020**, it will reach **44 zetta-bytes** or **44 trillion gigabytes**.²

90%

90% of the data in today's world has been created in the last **two years**.³

50
million

Over **50 million** scholarly articles are currently in existence.³

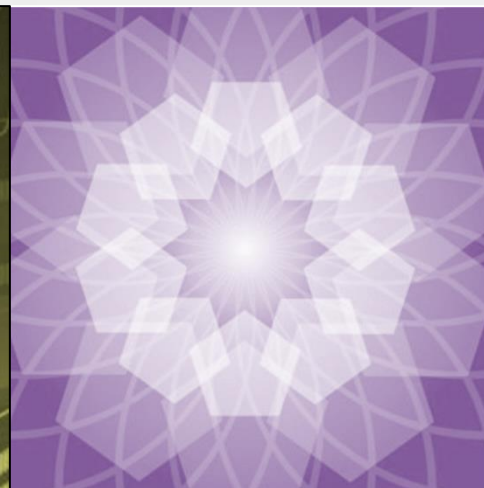
EVERY
13
SECONDS

2.4 million scholarly articles were published in 2013 or one article every **13 seconds**.⁴

\$300
BILLION
U.S.

The effective use of Big Data is worth more than **\$300 billion** per year to the U.S. healthcare sector alone.⁵

TEXT MINING – IT SEEMS SO EASY:



STAGE 1

Information
Retrieval

STAGE 2

Linguistic
Analysis:
Entity
Recognition

STAGE 3

Information
Extraction

STAGE 4

Data Mining
Knowledge
Discovery

BUT IT'S NOT...



STAGE 1

Where do I find data?



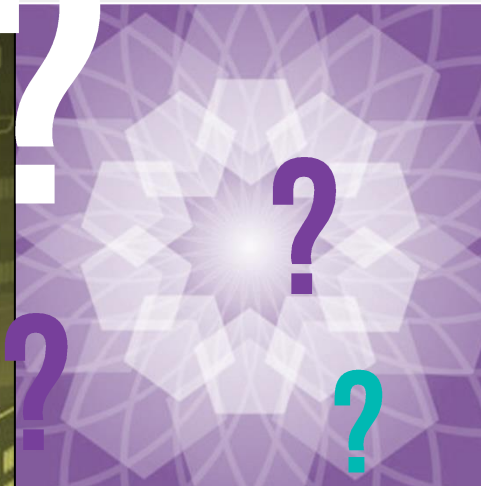
STAGE 2

How do I make my texts readable by machines?



STAGE 3

Which mining method to use?



STAGE 4

?

OPENMINTED WILL BUILD

Step 1: Registry service

Step 2: Workflow service

Step 3: Annotation & crowd-sourcing service

Platform for open...

..Mining information

Training, Q&A, guidelines,
Knowledge Hub on TDM
Hosted by FOSTER platform

..Mining tools

ALVIS, JCORE, U-COMPARE,
DKPro, ClearTK, Argo, GATE,
MetaShare

..Text and data

Scholarly Societies,
Publishers, OpenAIRE,
CORE, EuropePMC,
Repositories, Libraries



OPENMINTED BRINGS TOGETHER:

ACCESSIBLE CONTENT

Via standardised programmatic interfaces and access rules

DISCOVERABLE SERVICES

Easily discoverable text mining services and workflows which process, analyse and annotate text

EFFICIENT PROCESSING

Operate on public e-Infrastructures via standardized APIs

TDM COMMUNITIES

Different scientific communities have different challenges

VALUE ADDED APPS

Community-driven applications to illustrate the value of the infrastructure. Engage with industry.



LIBRARIES AS PART OF THE TDM COMMUNITY

We would like to work with you to...

- Make your data available for text and data mining
 - Born-digital
 - Digitalised data
- Provide you with the tools and knowledge to inform and encourage researchers to text and data mine
 - Where to find data
 - Where to find mining tools
 - Where to store and process the data
 - Where to find guidelines and training (both technical and legal)



BECOME INVOLVED

FOLLOW US ON TWITTER FOR THE LATEST UPDATES AND BLOGS

@OPENMINTED_EU

VISIT OUR WEBSITE

WWW.OPENMINTED.EU

SIGN UP FOR OUR NEWSLETTER

WWW.OPENMINTED.EU/NEWSLETTER



THANK YOU

@OPENMINTED_EU

WWW.OPENMINTED.EU



@OPENMINTED_EU

PROJECT FUNDED BY

