

# Machine Learning

A new essential tool for Knowledge Management

KMA - DC  
Januray 27, 2017

AI has always been a core part of information technology. It has already begun to catch up to the challenges facing Knowledge Management (KM), and will continue to accelerate. Can we take advantage of everything it has to offer? Perhaps more importantly, can we stay on the curve so that our insight grows along with it?

KM is about managing knowledge, but it's easy to get bogged down with data overload. It's helpful to think of AI as a tool that can automatically transform data into knowledge, and the many ways that might compliment KM.



## Big Data as a Challenge Facing KM

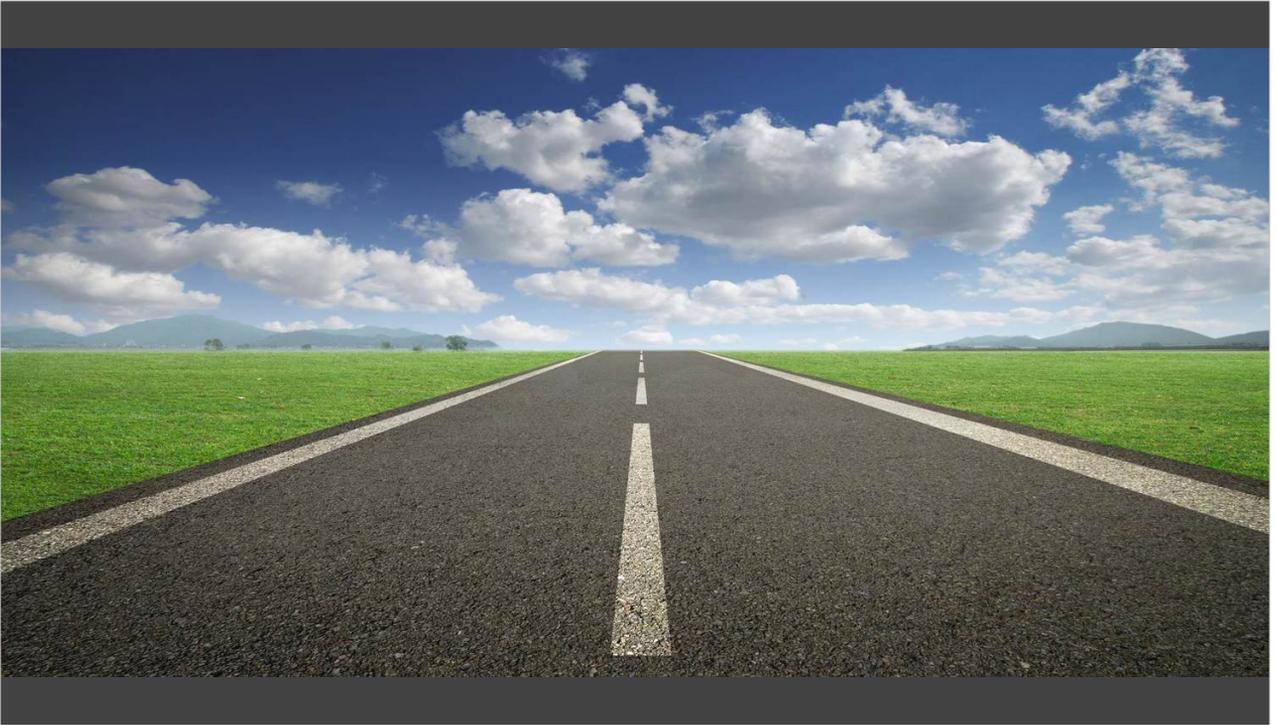
AI has now very much caught up with Big Data. You can easily apply advanced AI algorithms on massive data sets in an efficient way using tried and true Big Data algorithms (e.g. running topic extraction on a very large set of documents in real-time).



## Knowledge Sharing as a Challenge for KM

Knowledge sharing requires active effort on the part of the knowledge creator.

Automation and analytics can help create seamless sharing by offloading some of that responsibility.



Visibility and the Strategic Value of KM

AI + KM → Helps KM reach its full potential & grounds AI in real problems.

# Building Blocks

Let's talk about text baby



Building Blocks



Search



Taxonomy



Users



Knowledge Sharing

There is a lot of hype surrounding AI.

By understanding the technology a bit more, we can better navigate the terrain. At its core, AI means exciting changes in how things are being done, with interesting, and often unexpected applications.

# Demystifying the lingo

## Artificial Intelligence

### Tools

- Machine Learning
  - Neural Networks (deep learning)
  - Support Vectors
  - Probabilistic Graphs
  - Etc.
- Symbolic / Logic

### Fields

- Natural Language Processing
- Computer Vision

AI is the umbrella term for everything.

ML is the tool kit we use to solve problems in AI.

NLP and Computer Vision are fields in AI.

Recently, applications of Deep Learning (which is just a term for big or “deep” Neural Networks) have had excellent results in a variety of domains, but other algorithms are still important.

What’s new is understanding how to represent unstructured content (like images and text).

## Symbols

b a k e

b r o i l

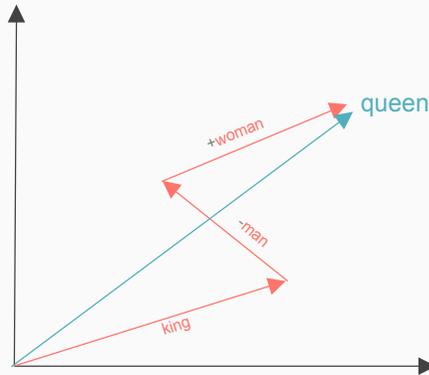
Words are made up of letters which are just symbols.

People obviously use words to convey ideas. For example, bake and broil are “close” words because the ideas are similar.

Up until only a few years ago, the computer only saw words as symbols. That meant bake and broil would only be considered “close” because they share the letter “b”.

## Vectors

$$\text{king} - \text{man} + \text{woman} \cong \text{queen}$$



Vectors are just numbers.

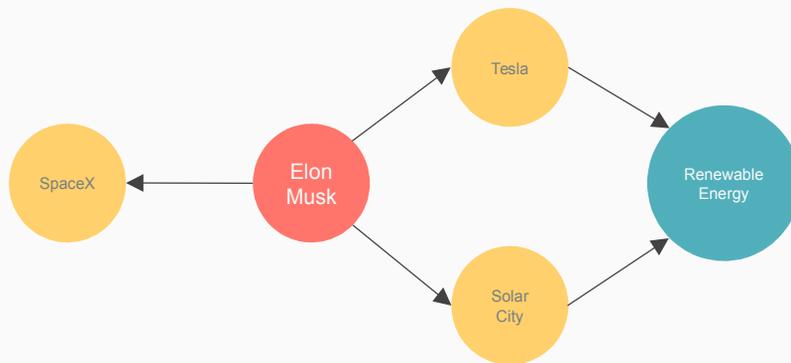
Recent techniques let us automatically assign vectors to words in a way that preserves meaning.

So a vector representation of a word – a bunch of numbers – is the meaning of that word. That means we can add and subtract these meanings

Vectors can also tell us when two words have similar meanings: because the vectors will be close to each other.

Now the computer can deal with words like ideas, instead of as just symbols.

# Elon Musk resigned yesterday.



## Link Text to Knowledge

Nowadays it's fairly straightforward to automatically assign labels to text based on the type of entity. For example, a computer could recognize that "Elon Musk" is a person based on the grammar of a sentence.

What's interesting now is we can link that reference to a body of knowledge about the entity. The more context we have, the richer the understanding of the text.

We've progressed well beyond the symbols that constitute our words.

“We are trying to replace symbols by vectors so we can replace logic by algebra.”

Yann LeCun

Rules are hard to create and harder to maintain.

Vectors are fluid and are learned from the content itself.

Instead of rules we now try to find the best “natural” representation as vectors.

Then we can manipulate those vectors to do the same thing we were trying to do with rules.

# Search

Beyond keywords



## Current Solutions



elastic



Open source  
Distributed  
Fault tolerant  
Excellent performance

But...

All are keyword based and filter based on document metadata.

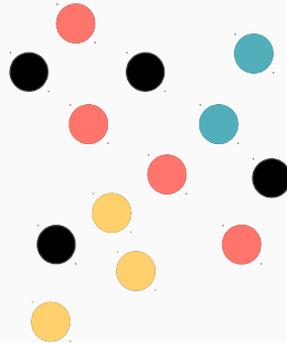
## Using vectors

fracking

hydraulic drilling

natural gas

cementing



### Enhanced Search

If someone is searching for “fracking”, wouldn’t it be nice to also surface documents that mention “hydraulic drilling”?

We can do that with vectors automatically (and at the option of the user).

## Using knowledge

French pharmaceutical company license

SEARCH

OSE Immunotherapeutics inks license option pact with Servier

US Army's planned Zika vax license to Sanofi raises nonprofit's ire

We can also use the properties of entities to filter.

To search, you must first define:

- 1) Where you're searching
- 2) What you're looking for

With text linked to knowledge we can filter where we're searching automatically.

Now imagine combining that with vectors!

## Advantages



Hands free



Keeps going

A lot of the value provided by AI is hands free.

Vector representations are automatically built from, and specific to, the text.

For example, in Finance the word “share” is close to “stock” or “equity”. In a regular context it would be closer to “collaborate”.

This context can benefit greatly from public or Open data sets, but that’s only half the equation. Internal data, or other information specific to your organization, must be integrated into your AI solution in order for it to provide real value.

Algorithms are also always learning and updating. That continuity is valuable, and maintaining it is easy.

# Taxonomy

Discover & Adapt



Building Blocks



Search



Taxonomy

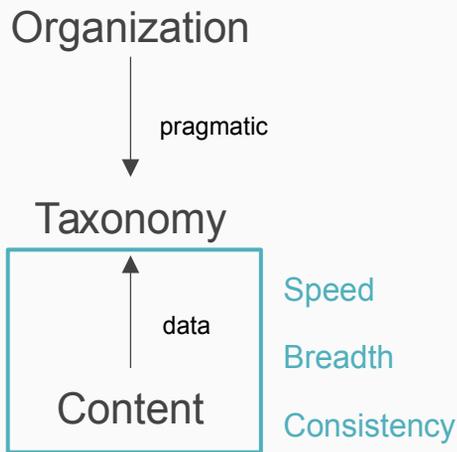


Content Management



Knowledge Sharing

## Creating



Taxonomy is a difficult task and requires a lot of careful attention.

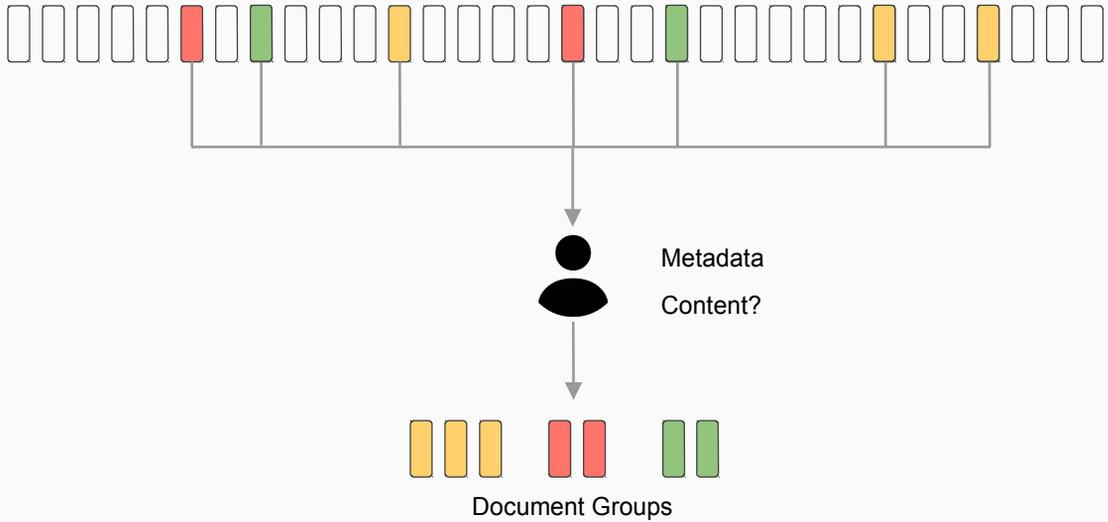
A lot of the thinking/planning that goes into creating a taxonomy is outside the scope of AI, or even technology.

AI can free you to be more creative by cutting down on the time it takes to evaluate taxonomies vs. content. This can enable you to test and iterate more rapidly.

Not only does that mean going faster, it also lets you explore more territory.

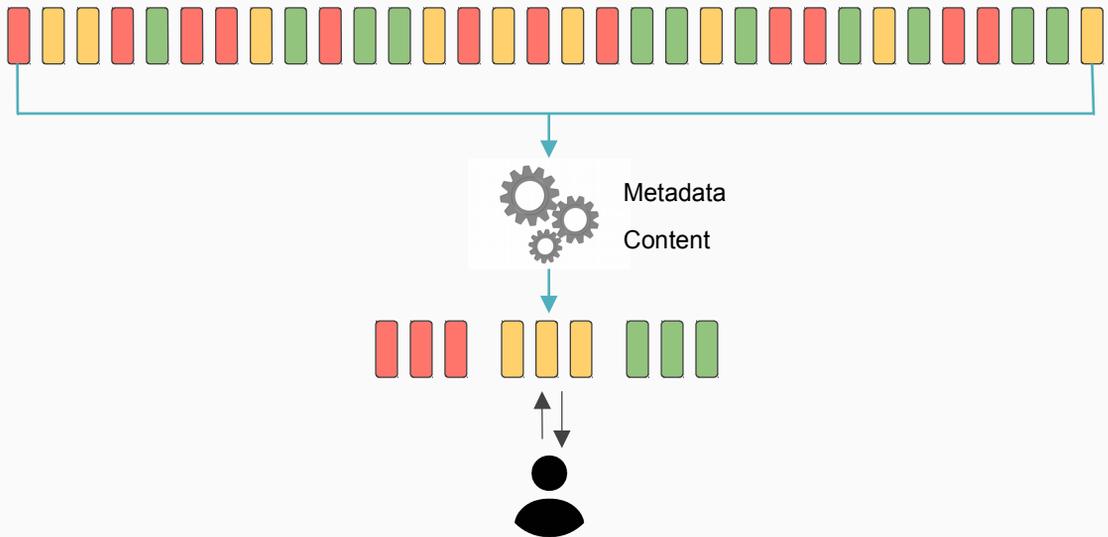
# Discovery

Content



The current method for creating a new taxonomy is to sub-sample (from a data perspective), possibly relying on content, but primarily on metadata.

# Discovery



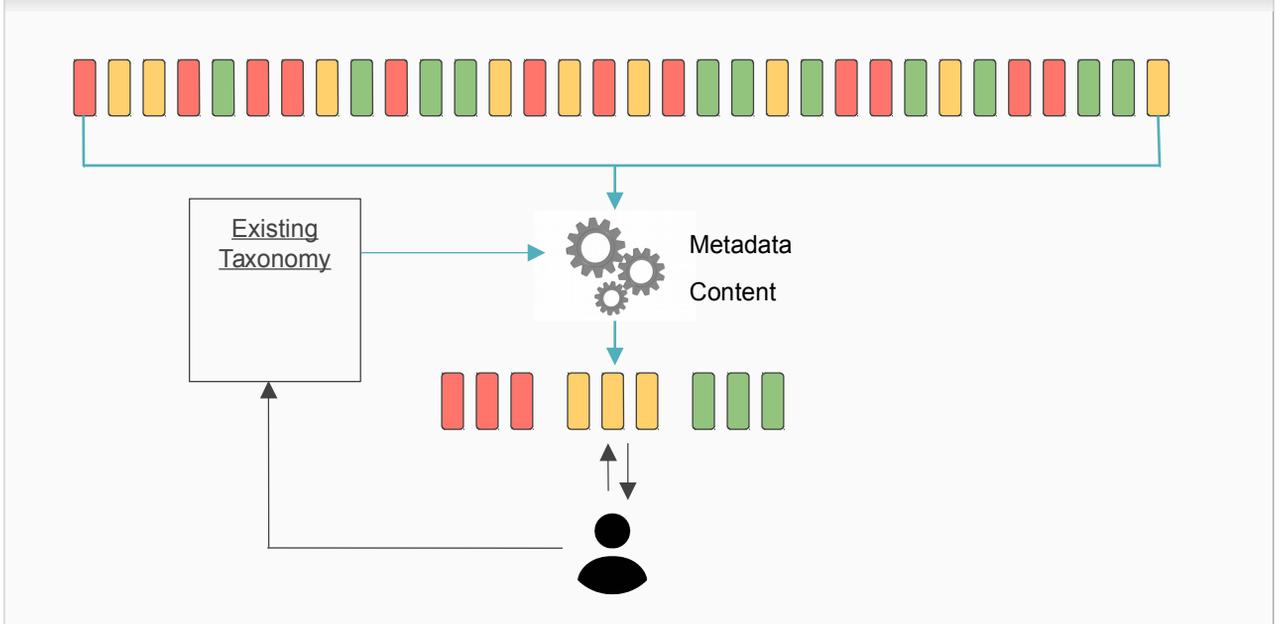
## Discovery

Group documents by analyzing the content of documents, not just the metadata.

Auto-tagging is powerful when using text linked to knowledge and vectors. It results in tags that are very rich in meaning and correspond nicely to each document (sometimes in surprising ways!).

This can also simplify concept hierarchies to collapse or expand related tag groups (e.g. “deforestation”, “water management”, “renewable energy” => “sustainability”) and helps automatically group the documents based on their content.

# Discovery

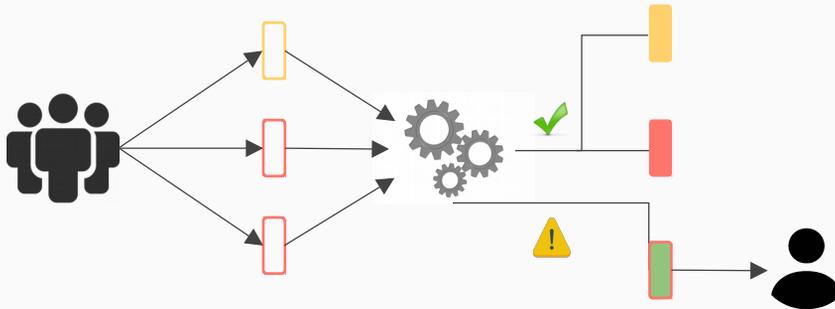


Integrate an existing taxonomy or build a taxonomy in tandem with the analysis.

An existing taxonomy may have short descriptions for each category. Using vectors and knowledge exploration we can look at each document and measure how “close” it is to each short description and propose a category for it based on the existing taxonomy. If it isn’t really close to any description, it can be placed in an “unassigned” category for evaluation.

The most important lesson here is the ability that AI provides us with to manipulate and understand text based on the meanings of the words and phrases, instead of just comparing symbols.

## Monitoring - mislabelling



Over time mislabelling reduces the value of a taxonomy as it becomes harder and harder to fix the bad tagging.

The same tech as described in the section on Discovery can be used to perform real-time audits of proposed labels.

It only gets better with time: we can compare new documents to documents already categorized and as our data grows, our results improve.

In the discovery phase we're trying to automatically group content; at this stage we already have the content grouped so the question is a bit easier to ask: does this new document fit with all these other documents?

## Strategic value



Detect mis-labelling



Adaptive taxonomy

Content may not fit existing labels. A category may need to be split.

Top down there may be organizational changes (new lines of business, new projects, re-organization of operating segments, etc.) that require changes to your taxonomy. How can you do that today?

The same tech we're talking about can also discover anomalies in classification systems when they get large enough.

These methods taken together mean you can have a much more flexible/organic/adaptive taxonomy.

# Users

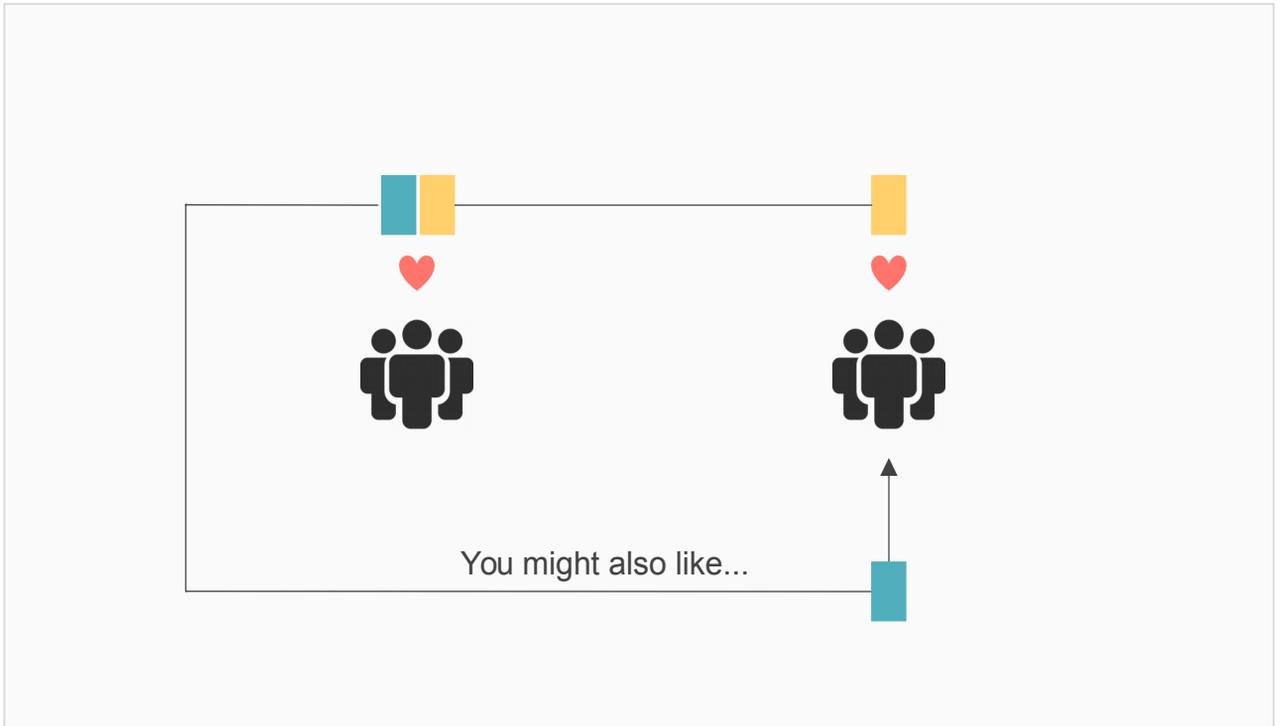
I'll tell you what I want



Users means both internal (employees) and external (customers) – whoever is consuming your content.

“Content flooding” is a problem. How do you give users relevant content?

How do Amazon or Netflix do it?



## Collaborative Filtering

Group A likes Y & T, Group B likes Y.

Both groups like Y.

Maybe Group B likes T as well?

This requires a lot of data to work well.

Most people don't have the volume of users Amazon and Netflix do, so how do you do it?

## I'll tell you what I want

What I really  
really want...



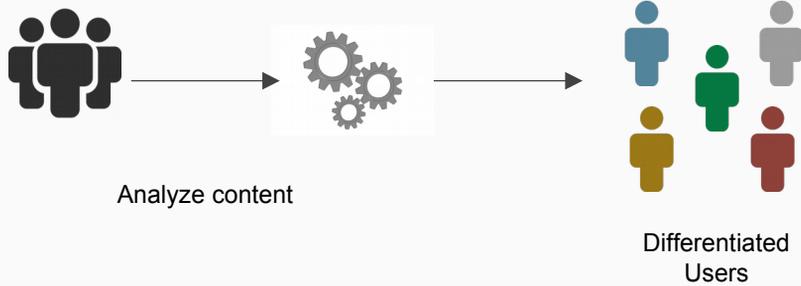
- ✓ Using your content
- ✓ Creating new content
- ✓ Attending events or conferences
- ✓ Interacting with other users
- ✓ User profiles

Users are always telling you what they want, we just have to figure out how to understand what they're saying.

With text understanding you can analyze this data quickly and efficiently.

It's the same engine that analyzes your content, but now pointed towards content related to users.

## Mapping users



Take undifferentiated users → Analyze content related to them → Create differentiated users.

This must adapt to users' behavior so this has to be a real-time system:

- Initial analysis helps you bootstrap your guess (i.e. take a first shot at what users might like)
- Need to keep adapting, but bootstrapping helps you bridge the data gap

# Consumption



Serve content



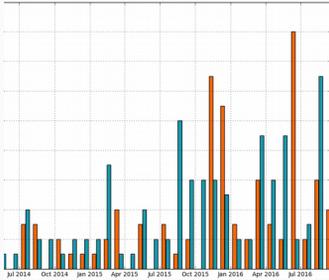
Find content

Now you can create a much better experience for users via push and pull (e.g. serving customized content on your site or ranking relevance in search respectively).

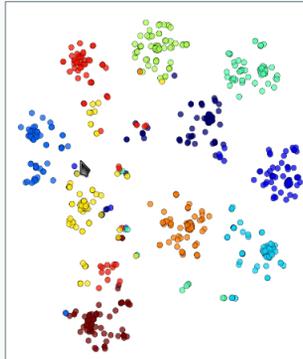
The models are also learning and start with good guesses.

Better guesses → more involvement → more data → more learning → better guesses → ...

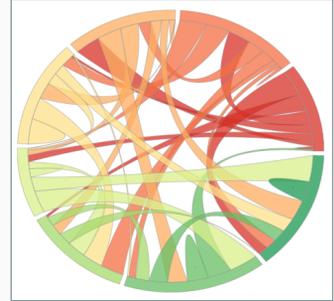
# Analytics



Trends



Groups



Networks

Very exciting because analysis can now create meaningful insights and help drive strategy.

## Examples

What people are interested in

What people are talking about

Grouping users

Changes over time

Connect conversations between different users

Internal and external

# Knowledge Sharing

Automation and analytics



Building Blocks



Search



Taxonomy



Users



Knowledge Sharing

## Recap

- 1) Vectors and knowledge
- 2) Search, moving beyond keywords and towards understanding and intuitive search
- 3) Taxonomy and tagging your content automatically, making it organic and fluid and adaptive
- 4) How this relates to users: now we can understand them either based on content they view or even better, what they create, their profiles, comments etc.
- 5) Now we're going to talk about knowledge sharing and support

Employees are a great knowledge asset.

## Bottlenecks



Too many requests



Out of date records

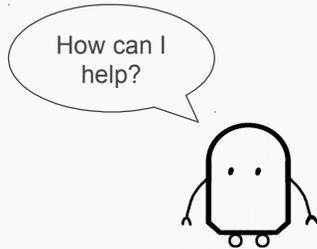
Requests can be internal or external → basically people asking questions that need answers.

Bottlenecks occur because too few people are available to answer those questions, whether it's help desk tickets desk or corporate yellow pages (to look up subject matter experts).

In the case of yellow pages, they get outdated too.

Either way phone/email or open a ticket to ask questions creates bottlenecks and constrains the flow of info.

## Bots



✓ Access anywhere

✓ 24/7

✓ Volume

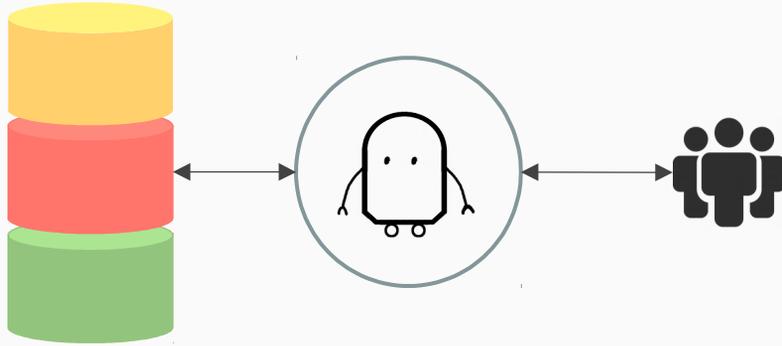
✓ Fast

✓ Escalate

How about a bot?

If properly executed, bots can at least relieve some pressure on the bottlenecks and create more fluid engagement. Already we can answer many questions using a bot. Very soon we'll be able to understand (and therefore answer) a lot more, ultimately creating a Siri for your business.

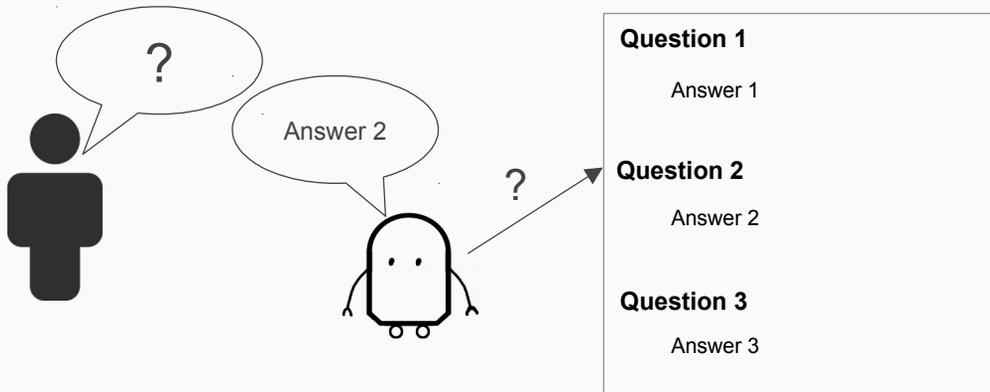
## Privacy and security



Security is a real concern.

We can create security “bubbles” around the bot both for data access and for user access to bot functionality.

# I have your answer

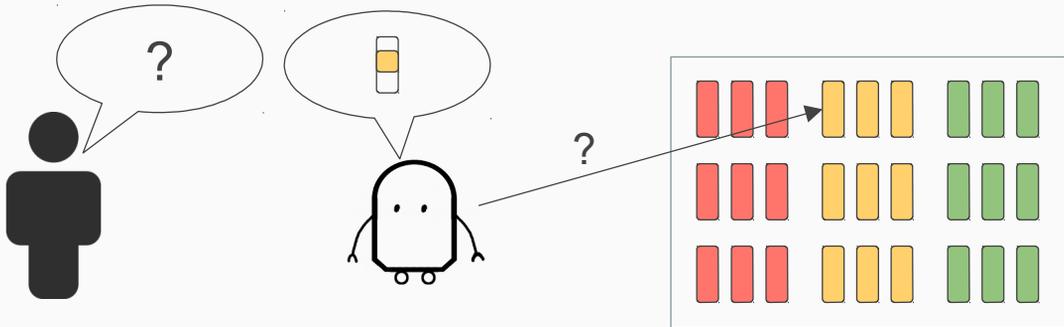


Question to question mapping based on stored answers (e.g. FAQs).

Easy to implement.

Get all the benefits of the always-available bot interface, plus handling many variations in the way questions are posed.

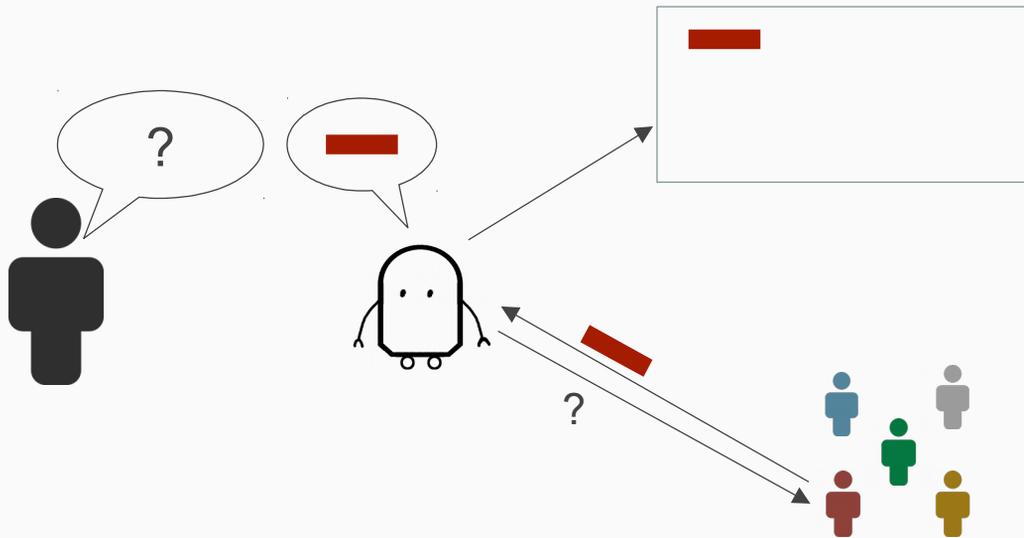
## I'll find your answer



The next layer is for the bot to seek out answers itself in your content.

Structured data (databases etc.) and unstructured data (docs/text) can both be integrated with the bot so it can go out and find the answer if it doesn't already have it.

## I know who has your answer

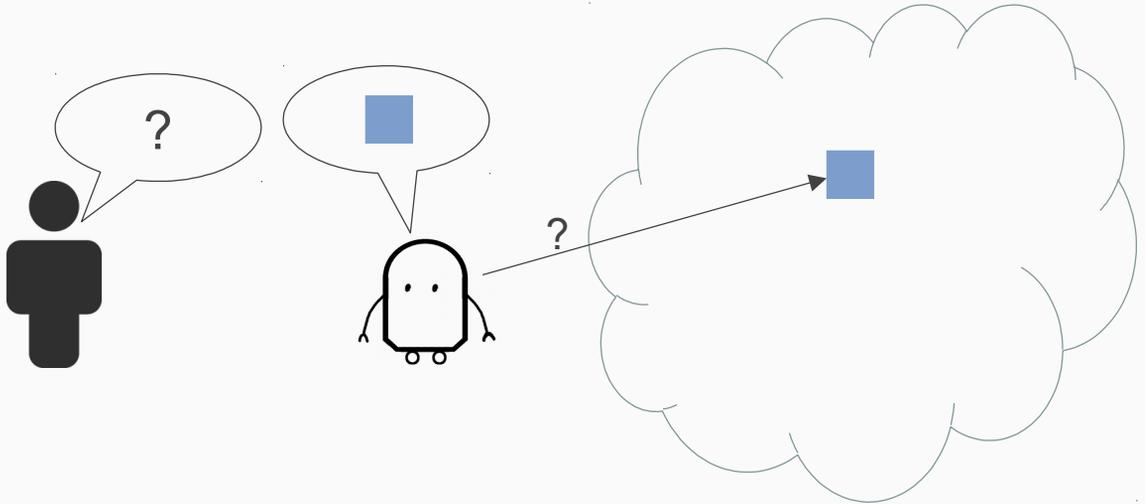


Because we can map users to content (from earlier section) what if we also mined employee CVs/resumes, job descriptions, profiles, or even emails to find subject matter experts who can answer the question?

Still requires effort from the employee BUT...only once since we store the answer after the first time.

This is automated tacit knowledge accumulation → like automatically building a knowledge base through organic/natural interactions.

## It's out there somewhere

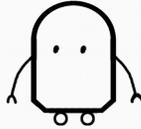


Last, we can start connecting to the Internet to get answers.

News, open data, linked data, unstructured content on the web.

This is also very exciting because now we're basically pulling in info from outside an organization, but only to the extent that it's relevant to answering questions inside the organization. Very powerful way to manage knowledge.

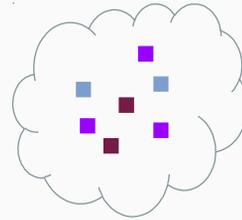
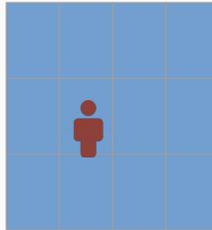
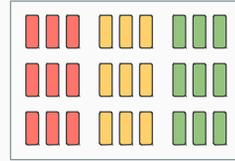
# Ask me anything



Question 1  
Answer 1

Question 2  
Answer 2

Question 3  
Answer 3



Through automation you can significantly reduce work load from manual, repetitive tasks and the ask on people in the organization to support sharing

But, we still need to manage/administer and then more importantly, identify what questions need to be answered and maybe even why some questions don't yet have answers but are being asked a lot!

I'd like to suggest that not only does the work become easier but the nature of the work becomes more about analytics → insights from interactions etc.

In some ways, this bot is a great embodiment of AI + KM.

Thank you  
Hedeer El Showk  
hedeer@lore.ai



Building Blocks



Search



Taxonomy



Users



Knowledge Sharing

We've just scratched the surface of problems and possible solutions.

By applying the correct AI technology, you can address many problems KM is facing in really amazing ways.

We'd love hear from you, AI community needs experts to connect solutions to real problems.