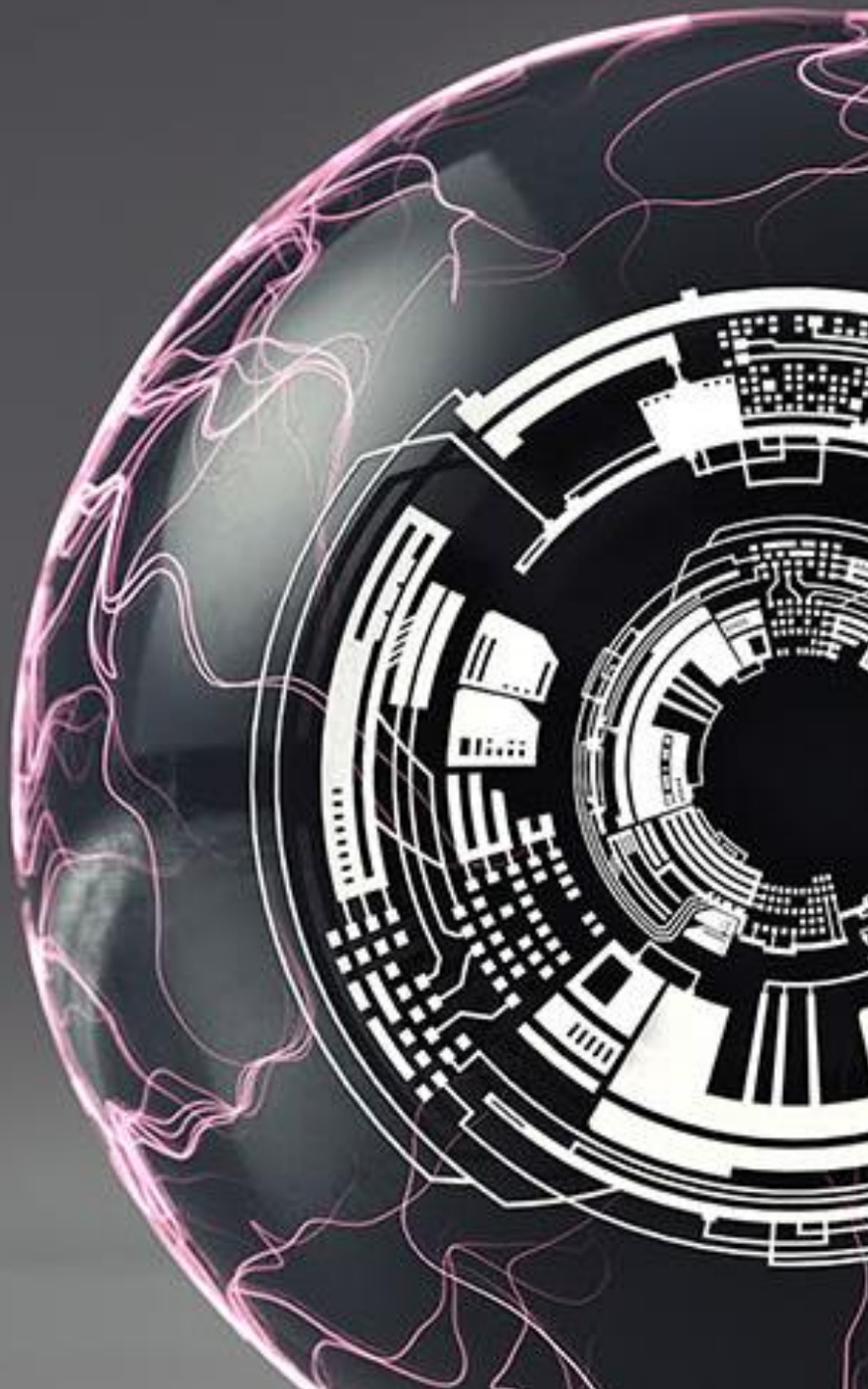
Microsoft

Semantic Scholar

Oren Etzioni, CEO
Allen Institute for AI
(Professor, CSE, University of Washington)



Allen Institute for AI (AI2)

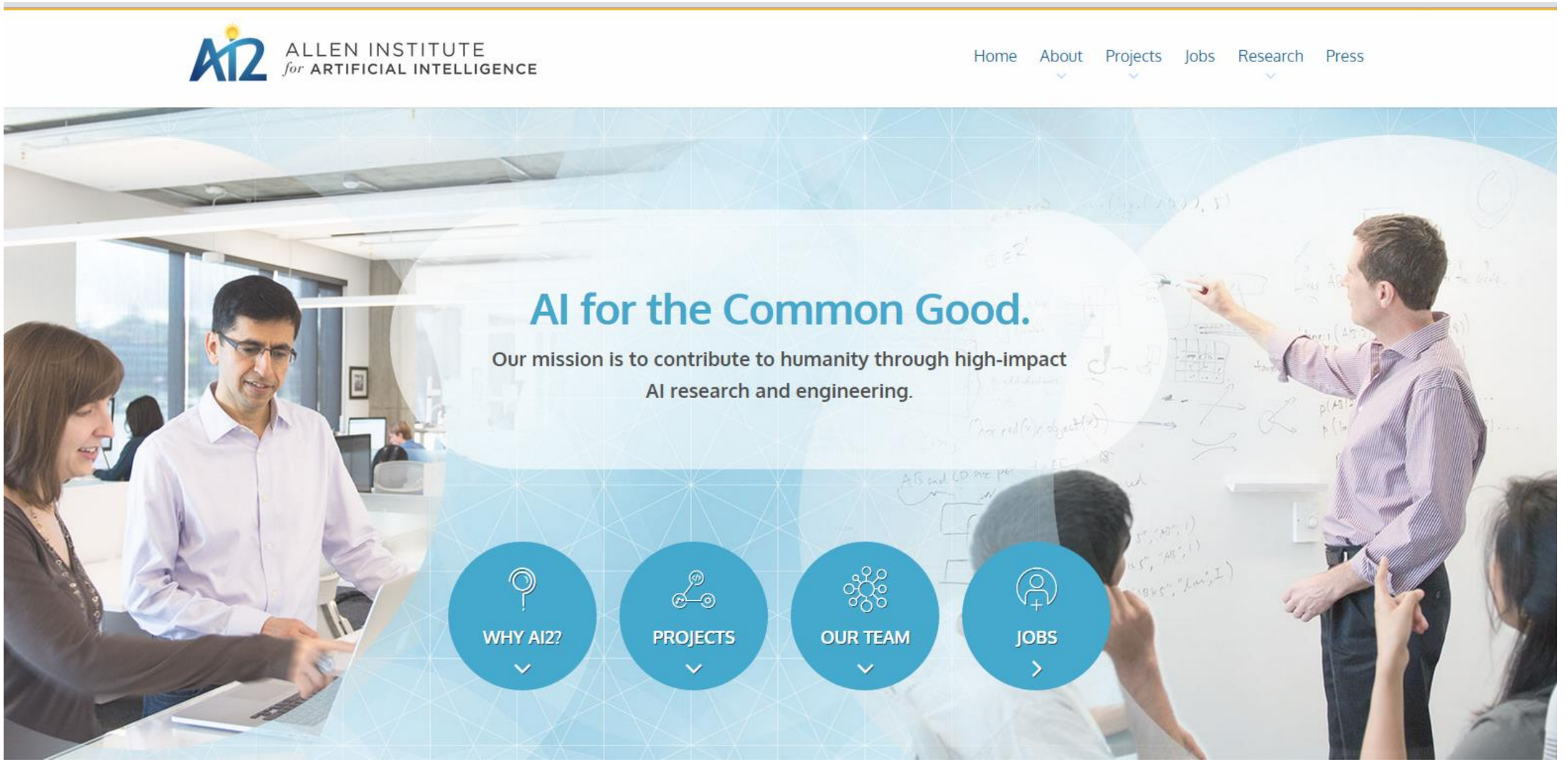
Founded by Paul Allen in 2014
Now 60 People

AI2's Research Methodology

Research driven by Grand Challenge Problems

- **Ambitious goals:** with measurable milestones

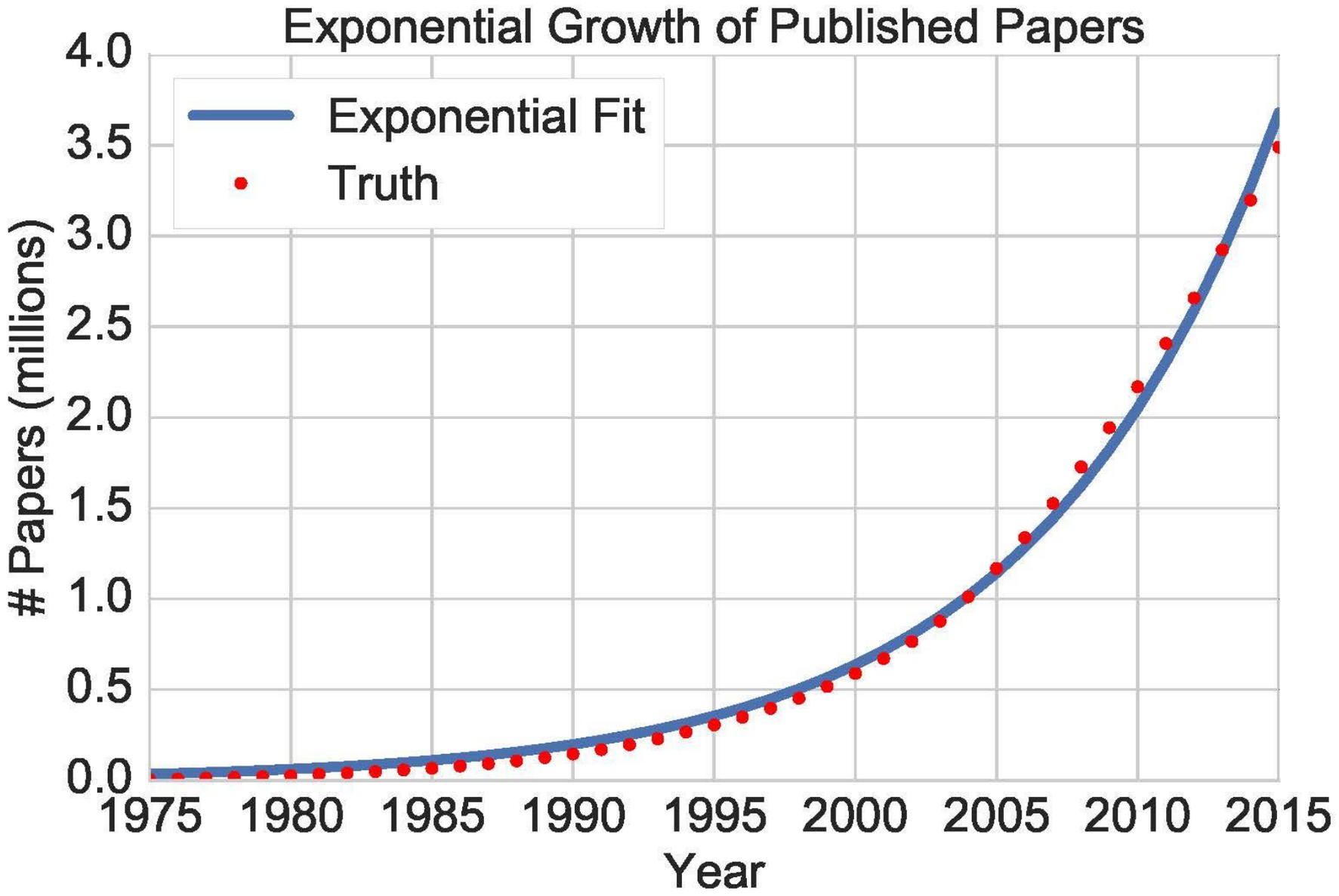
Focused on research tasks,
Agnostic on specific mechanisms


ALLEN INSTITUTE
for ARTIFICIAL INTELLIGENCE
Home
About
Projects
Jobs
Research
Press
AI for the Common Good.
Our mission is to contribute to humanity through high-impact
AI research and engineering.
WHY AI2?
PROJECTS
OUR TEAM
JOBS

Moor

The

very



Source

Articles
Case law
My library

Any time
Since 2016
Since 2015
Since 2012
Custom range...

Sort by relevance
Sort by date

include patents
include citations

Create alert

Learning in science: A comparison of **deep** and surface approaches
C Chin, DE Brown - Journal of research in science teaching, 2000 - Wiley Online Library
Abstract The purpose of this study was to explore in greater depth what has been called by previous researchers, a **deep** versus surface approach to **learning** science. Six Grade 8 students judged as typically using **learning** approaches ranging from **deep** to surface were ...
Cited by 378 Related articles All 5 versions Cite Save

[HTML] Why does unsupervised pre-training help **deep learning**? [HTML] from jmlr.org
D Erhan, Y Bengio, A Courville, PA Manzagol... - ... of Machine **Learning** ..., 2010 - jmlr.org
Abstract Much recent research has been devoted to **learning** algorithms for **deep** architectures such as **Deep** Belief Networks and stacks of auto-encoder variants, with impressive results obtained in several areas, mostly on vision and language data sets. ...
Cited by 626 Related articles All 27 versions Cite Save

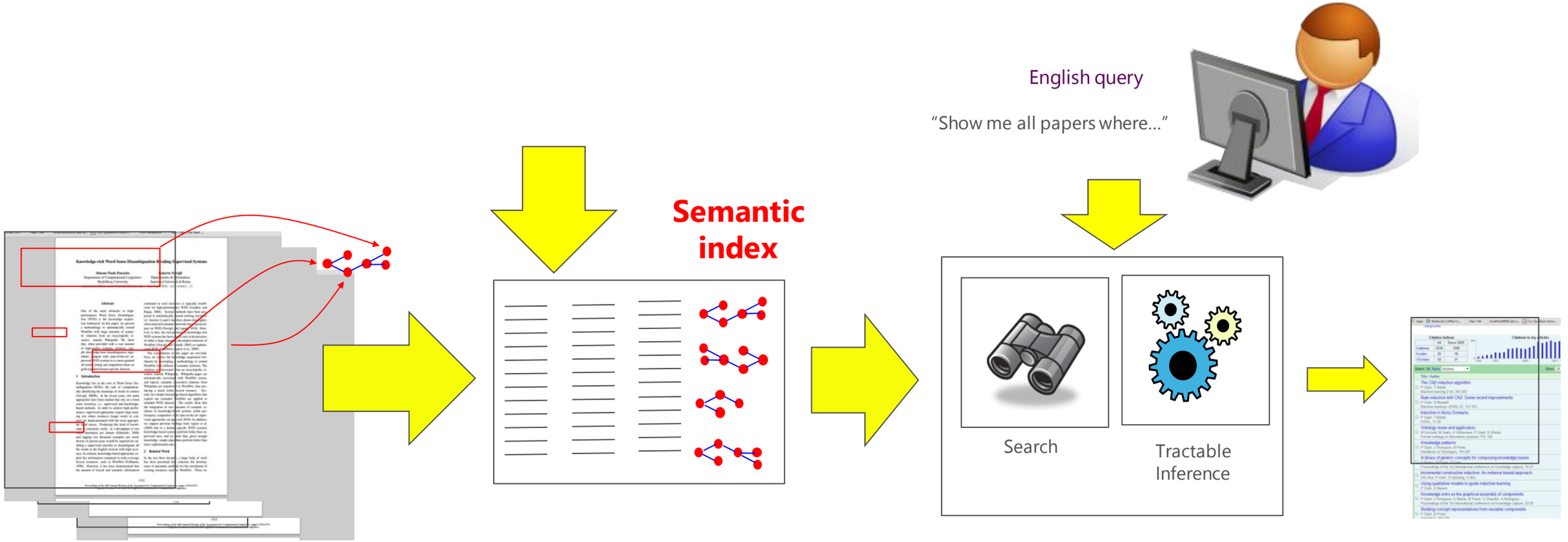
[PDF] Multimodal **deep learning** [PDF] from wustl.edu
J Ngiam, A Khosla, M Kim, J Nam... - ... machine **learning** (..., 2011 - machinelearning.wustl.edu
Abstract **Deep** networks have been successfully applied to unsupervised feature **learning** for single modalities (eg, text, images or audio). In this work, we propose a novel application of **deep** networks to learn features over multiple modalities. We present a series of tasks for ...
Cited by 433 Related articles All 29 versions Cite Save More

Deep learning [PDF] from bioinfo.org.cn
Y LeCun, Y Bengio, G Hinton - Nature, 2015 - nature.com
Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction. These methods have dramatically improved the state-of-the-art in speech recognition, visual object ...
Cited by 632 Related articles All 12 versions Cite Save

[HTML] Unsupervised feature **learning** for audio classification using convolutional **deep** belief networks [HTML] from nips.cc
H Lee, P Pham, Y Largman, AY Ng - Advances in neural information ..., 2009 - papers.nips.cc
Abstract In recent years, **deep learning** approaches have gained significant interest as a way

Comprehensive follow up on citations and references requires super-human diligence

Idea: Leverage AI to Combat Information Overload



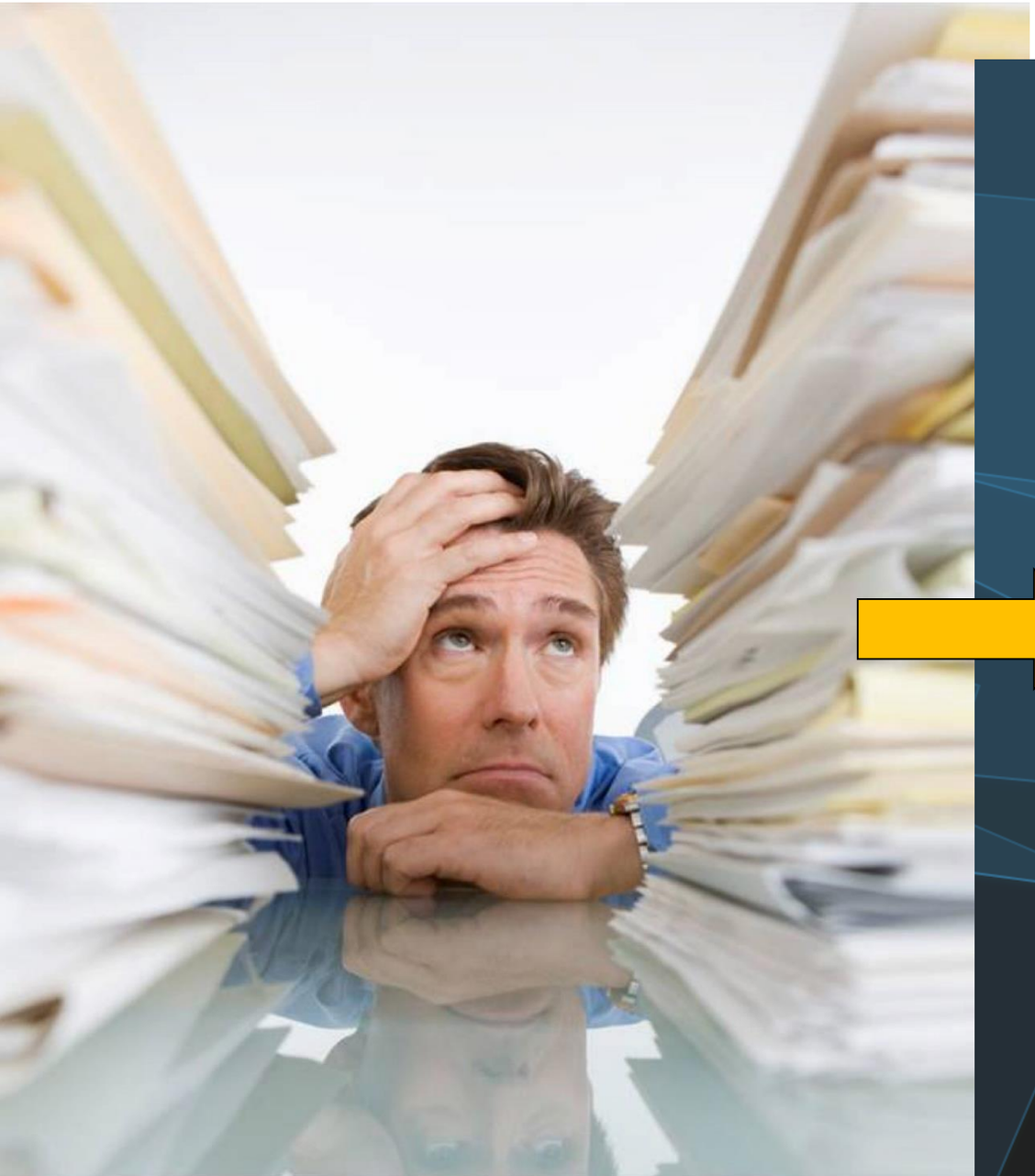
Machine Reading to the Rescue?

"The time is ripe for the AI community to set its sights on *Machine Reading*." (Etzioni, *et al. AAAI, 2006*)

Open Information Extraction:

- Arbitrary topics
- Minimizes labeled data

Semantic Scholar (Scientific Search)



Notable Collaborators

- **Douglas Downey,** Northwestern University
- **Marti Hearst**, UC Berkeley
- **Lee Giles** and the CiteSeer team
- Marcel Ackermann and the DBLP team
- **Jie Tang** and the AMiner team
- Alex Wade and Microsoft Academic Search

Unique Capabilities of Semantic Scholar

Facets to home in on results

- **Overview via key phrases**
- **Data sets used**

Influential citations/references

- Determined by learned classifier
- Ignores self citations

Home in on Figures, tables, and citation excerpts

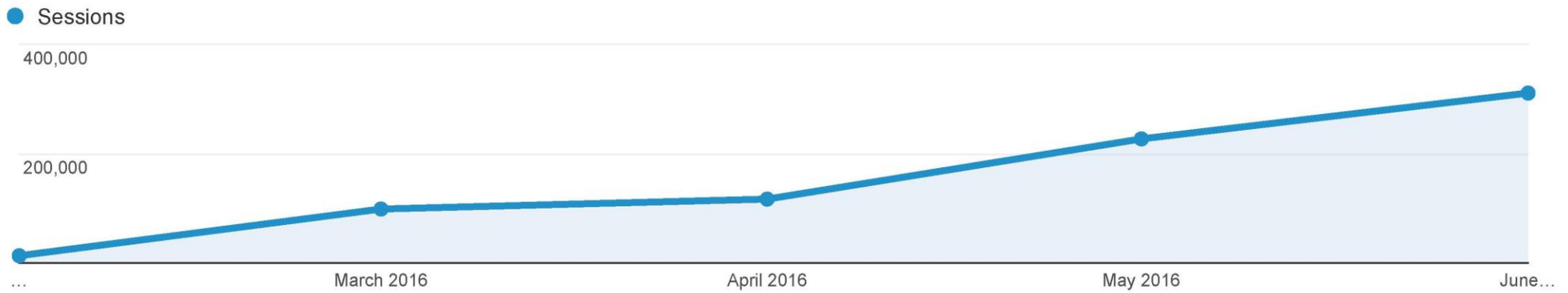
DEMO

Audience Overview

Feb 1, 2016 - Jun 30, 2016

All Users
100.00% Sessions

Overview



Sessions	Users	Pageviews
766,642	566,625	2,069,241

Pages / Session	Avg. Session Duration	Bounce Rate
2.70	00:01:39	60.20%

A Peak "Under the Hood"

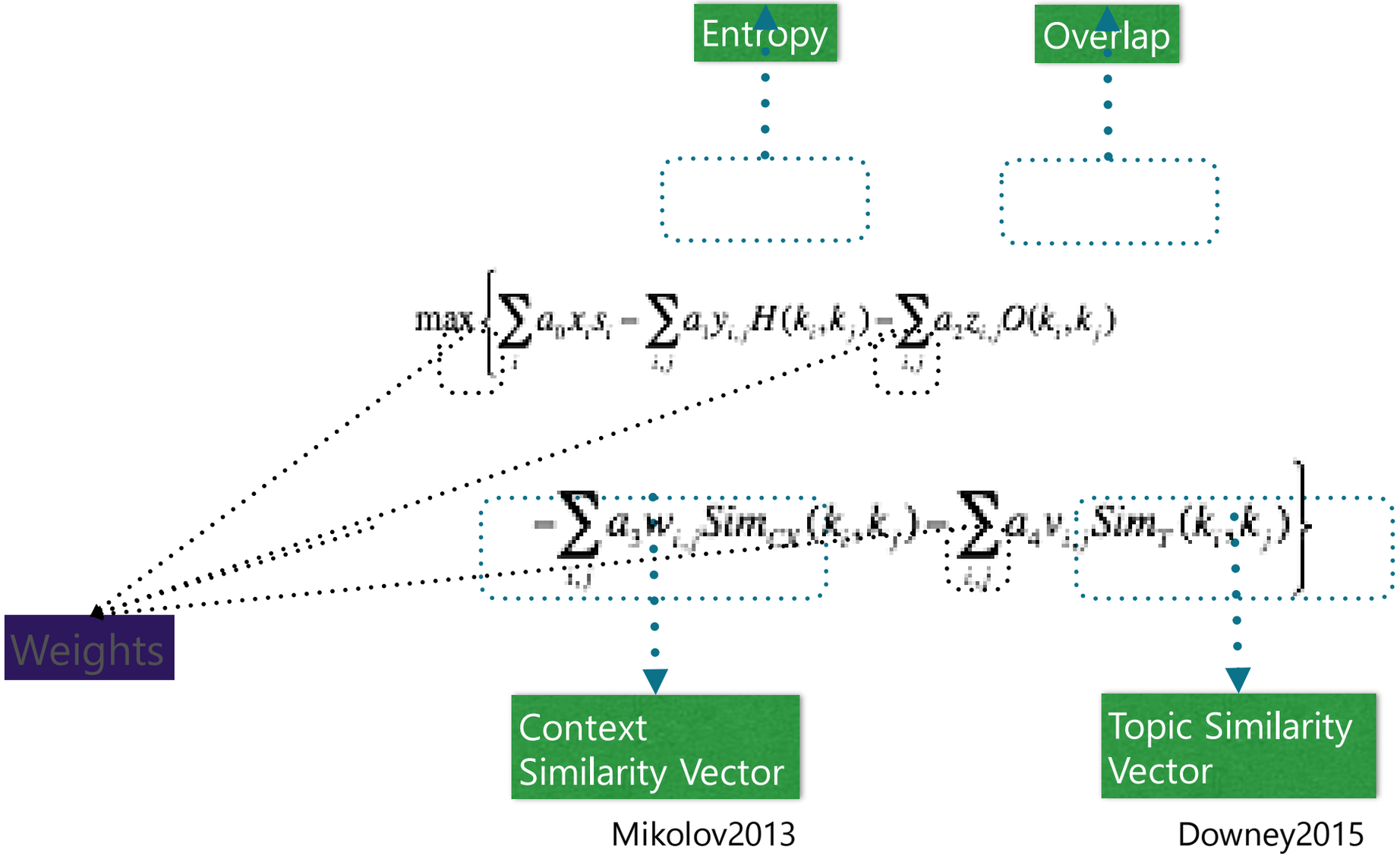
Automatic Key phrase extraction:

- **Goal**: Identify ***K*** phrases that *best* describe an article
- **Input**: article
- **Output**: A set of ***K*** phrases

What is *best* ?

- **Relevant**
- **Informative**
  - Must be non-redundant:
    “PMI” ≅ “Pointwise Mutual Information”
    “Open IE” ⊂ “Information Extraction”
- **Have Good Coverage**
  - Must occur frequently in the corpus

ILP Ranker Snapshot



Semantic Scholar Data--- available to the research community

Microsoft Research
Faculty Summit 2016

Can we Predict Citation Rates in the Future?

Experiment: train a variety of ML methods on paper & author citation rates.

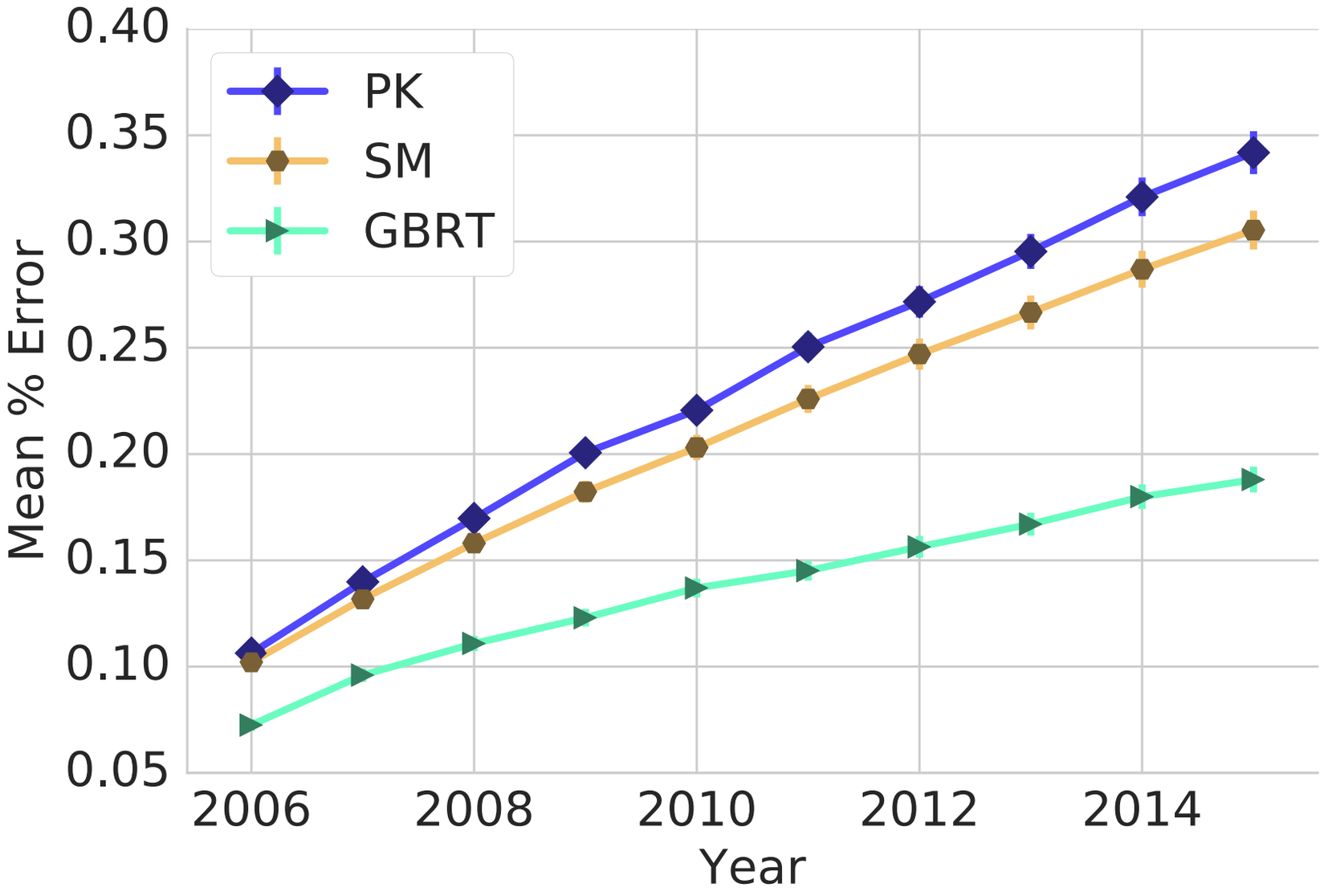
Training set: 1985-2005

Test set: 2006-2015

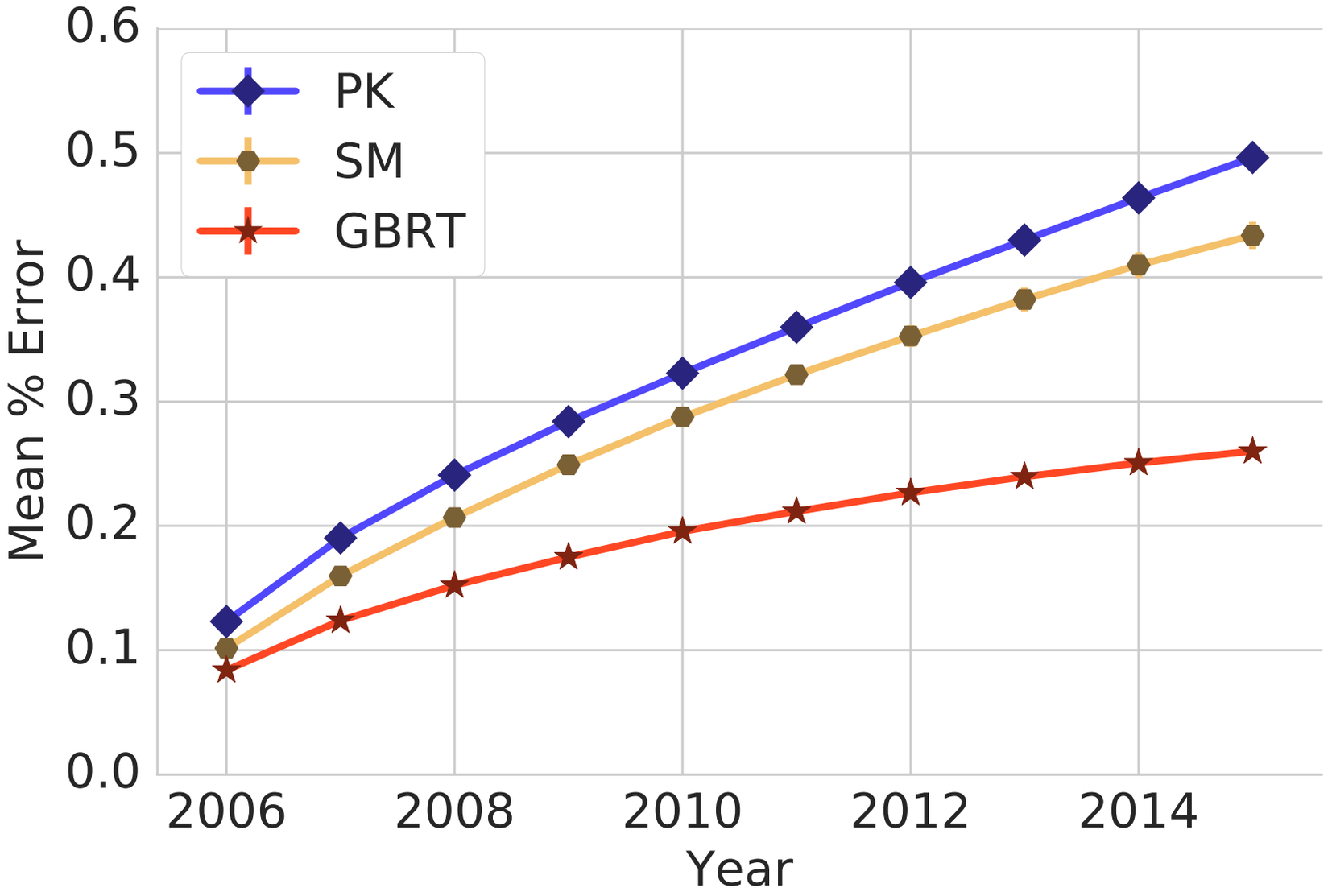
Metric: mean error rate

Joint work with Luca Weihs (UW Statistics)

Author h-index Prediction



Paper Citation Prediction



Semantic Scholar Future Work

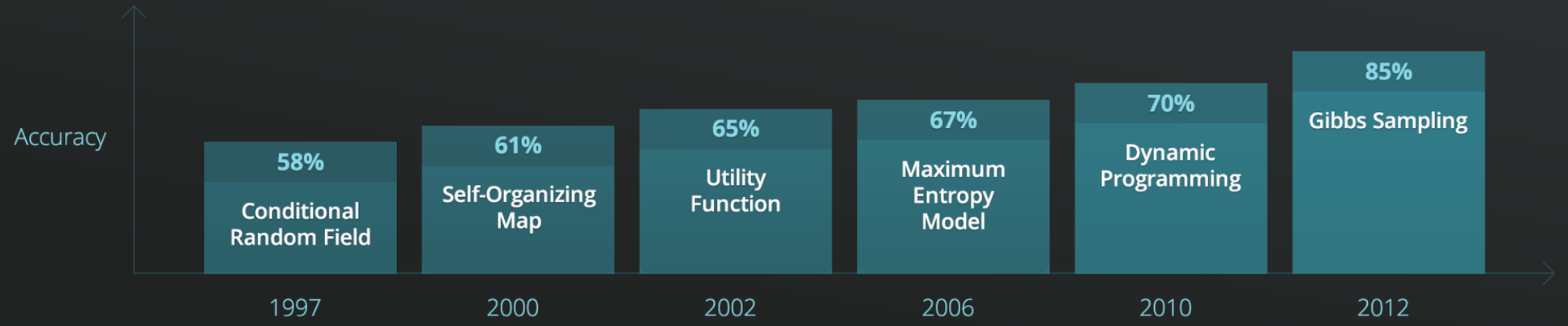
Semantic Scholar

information extraction

4

Information Extraction

Progress Summary



FILTER RESULTS

CLASSIFICATION

- Survey
- Experimental
- Theoretical
- Software

YEAR

6,962 results

Sort by: Relevance

Incorporating Non-Local **Information** Into **Information Extraction** Systems By Gibbs Sampling

Jenny Rose Finkel, Trond Grenager, Christopher D. Manning / ACL / 2005

Cited by 166 / Abstract / View PDF / Add to reading list

structure that is prevalent in language use. We show how to solve this dilemma with Gibbs sampling **information extraction** task. We show 10 runs of Gibbs sampling in the same CRF...

On-Demand Information Extraction

Satoshi Sekine / ACL / 2006

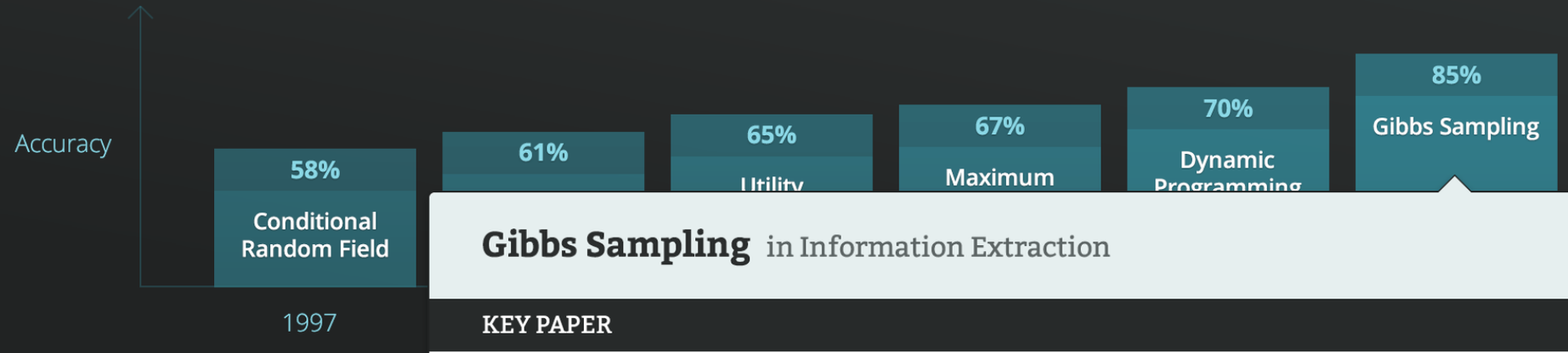
Semantic Scholar

information extraction

4

Information Extraction

Progress Summary



Gibbs Sampling in Information Extraction

KEY PAPER

Incorporating Non-Local **Information** Into **Information Extraction** Systems By **Gibbs Sampling**

Jenny Rose Finkel, Trond Grenager, Christopher D. Manning / ACL / 2012

An illustration of the effectiveness of **Gibbs sampling**, compared to Viterbi inference, for the two tasks addressed in theis paper: the CoNLL named entity recognition task **which returned an accuracy rate of 85.54%**, and the CMU Seminar Announcements **information extraction** task. We show 10 runs of **Gibbs sampling** in the same CRF model that was used for Viterbi. For each run the sampler was initialized to a random sequence, and used a linear annealing schedule that sampled the complete sequence 1000 times. CoNLL performance is measured as per-entity, and CMU Seminar. Announcements performance is measured as per-token.

FILTER RESULTS

CLASSIFICATION

- Survey
- Experimental
- Theoretical
- Software

YEAR

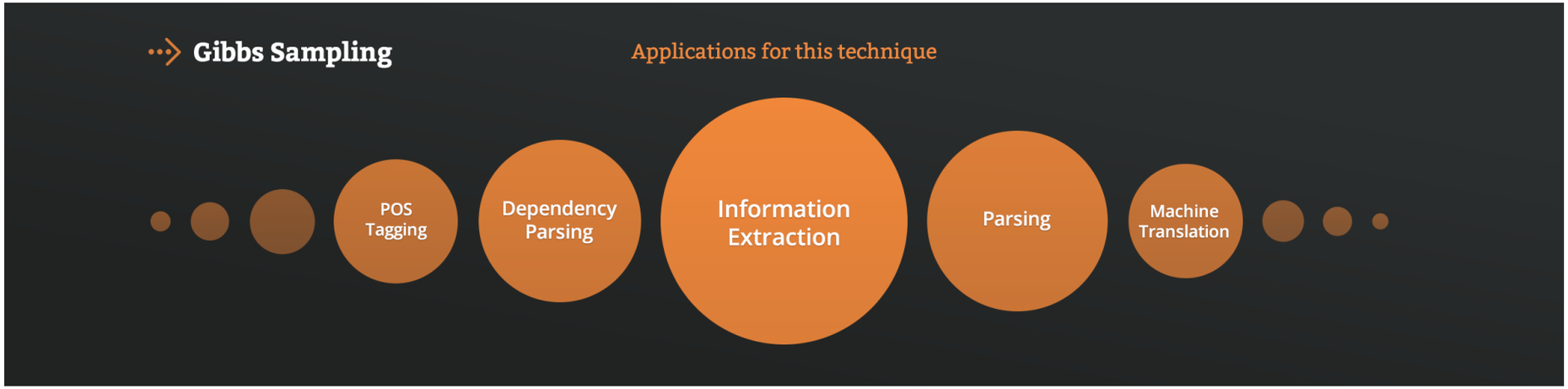
On-Demand Information Extraction

Satoshi Sekine / ACL / 2006

Semantic Scholar

gibbs sampling

4



FILTER RESULTS

CLASSIFICATION

Survey
Experimental
Theoretical
Software

YEAR

yyyy to yyyy

VENUES (15)

ACL

Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data

EMNLP

429 results

Sort by: Relevance

Incorporating Non-Local **Information** Into **Information Extraction** Systems By Gibbs Sampling

Jenny Rose Finkel, Trond Grenager, Christopher D. Manning / ACL / 2012

Cited by 166 / Abstract / View PDF / Add to reading list

structure that is prevalent in language use. We show how to solve this dilemma with Gibbs sampling information extraction task. We show 10 runs of Gibbs sampling in the same CRF...

Not-So-Latent Dirichlet Allocation: Collapsed **Gibbs Sampling** Using Human Judgments

Jonathan Chang / Proceedings of the NAACL HLT 2010 Workshop on Creating Speech ... / 2010

Cited by 1 / Abstract / View PDF / Add to reading list

Probabilistic topic models are a popular tool for the unsupervised analysis of text, providing both ... and cluster that annotation. This task simulates the **sampling** step of the collapsed **Gibbs** sampler

Sampling Alignment Structure under a Bayesian Translation Model

John DeNero, Alexandre Bouchard-Côté, Dan Klein / EMNMP / 2008

Cited by 31 / Abstract / View PDF / Add to reading list

We describe the first tractable **Gibbs sampling** procedure for estimating phrase pair frequencies

Semantic Scholar

gibbs sampling

4

Gibbs Sampling

Applications for this technique

POS Tagging

Dependency Parsing

Information Extraction

Parsing

Machine Translation

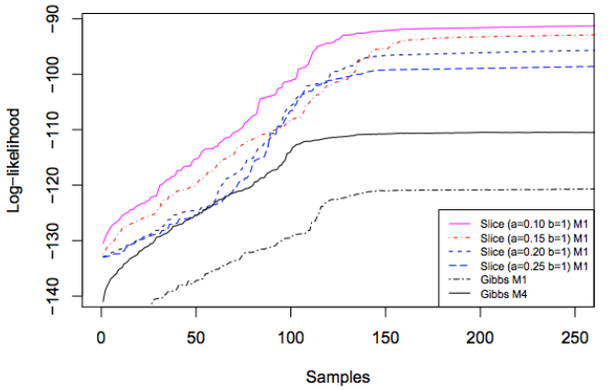
Dependency Parsing using Gibbs Sampling

KEY PAPER

Unsupervised Dependency Parsing using Reducibility and Fertility features

David Marecek, Zdeněk Zabokrtsky / NAACL / 2012

Inference	CoNLL	Seminars
Viterbi	85.51	91.85
Gibbs	85.54	91.85
Sampling	85.51	91.85
	85.49	91.85
	85.51	91.85
	85.51	91.85
	85.51	91.85
	85.51	91.85
	85.51	91.85
	85.51	91.85
	85.51	91.86
Mean	85.51	91.85
Std. Dev.	0.01	0.004



FILTER RE

CLASSIFIC

Surve

Exper

Theor

Softw

YEAR

Sort by: Relevance

Into **Information Extraction**

g / ACL / 2012

to solve this dilemma with Gibbs sampling
mpling in the same CRF...

lapsed **Gibbs Sampling** Using

Eric Horvitz

"It's the **absence** of AI technologies that is **already** killing people through errors."

The Semantic Scholar Vision

"What if a cure for an intractable cancer is hidden within the tedious reports on thousands of clinical studies? …AI readers will be able to connect the dots between disparate studies to identify novel hypotheses and suggest experiments which would otherwise be missed. **AI-based discovery engines will help to find the answers to science's thorniest problems** and ultimately revolutionize science."

Allen Institute for Artificial Intelligence
Wired Magazine, September, 2015