



#haystackconf EU keynote
Doug Turnbull

SOLR-8542



Solr / SOLR-8542

Integrate Learning to Rank into Solr

Details

Type:	+ New Feature	Status:	RESOLVED
Priority:	^ Major	Resolution:	Fixed
Affects Version/s:	None	Fix Version/s:	6.4, 7.0
Component/s:	None		
Labels:	None		

Description

This is a ticket to integrate learning to rank machine learning models into Solr. Solr Learning to Rank (LTR) provides a way for you to extract features directly inside Solr for use in training a machine learned model. You can then deploy that model to Solr and use it to rerank your top X search results. This concept was previously [presented by the authors at Lucene/Solr Revolution 2015](#).

We need to step into our time machines



Remember 2015?



What I think of when I hear 2015...



LUCENE / SOLR
REVOLUTION / 2015

Learning to Rank in Solr

- **Michael Nilsson**, Bloomberg LP
- **Diego Ceccarelli**, Bloomberg LP

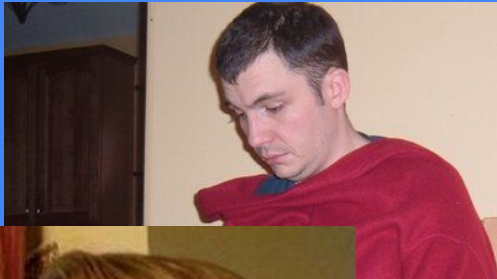
<https://www.youtube.com/watch?v=M7BKwJoh96s>

The image is a YouTube video thumbnail with a dark background and diagonal stripes. It features the text 'LUCENE / SOLR REVOLUTION / 2015' at the top left, a title 'Learning to Rank in Solr' in the center, and two speakers listed below: Michael Nilsson and Diego Ceccarelli, both from Bloomberg LP. A URL is provided at the bottom. A small 'i' icon is in the top right corner, and a red YouTube logo is in the bottom right corner.

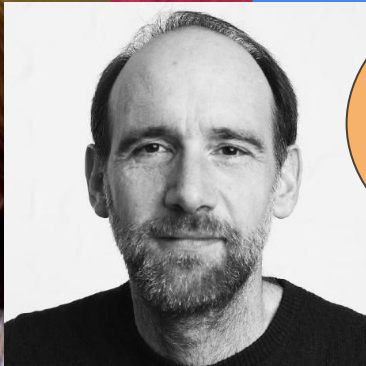


Solr LTR: a sea-change for
relevance community

pre 2015 life:



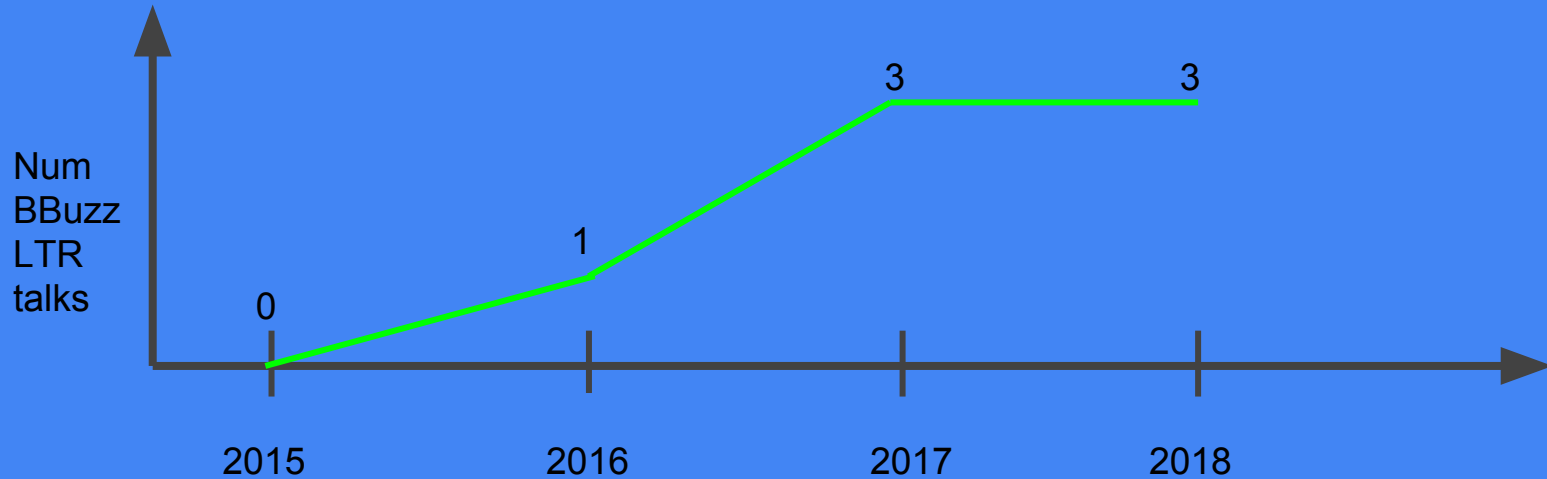
We want to focus
our consulting on
search relevance!



Nobody **gets** search
relevance, they just
want a pretty
single-page search UI
with some facets!



Number of LTR talks (bbuzz)

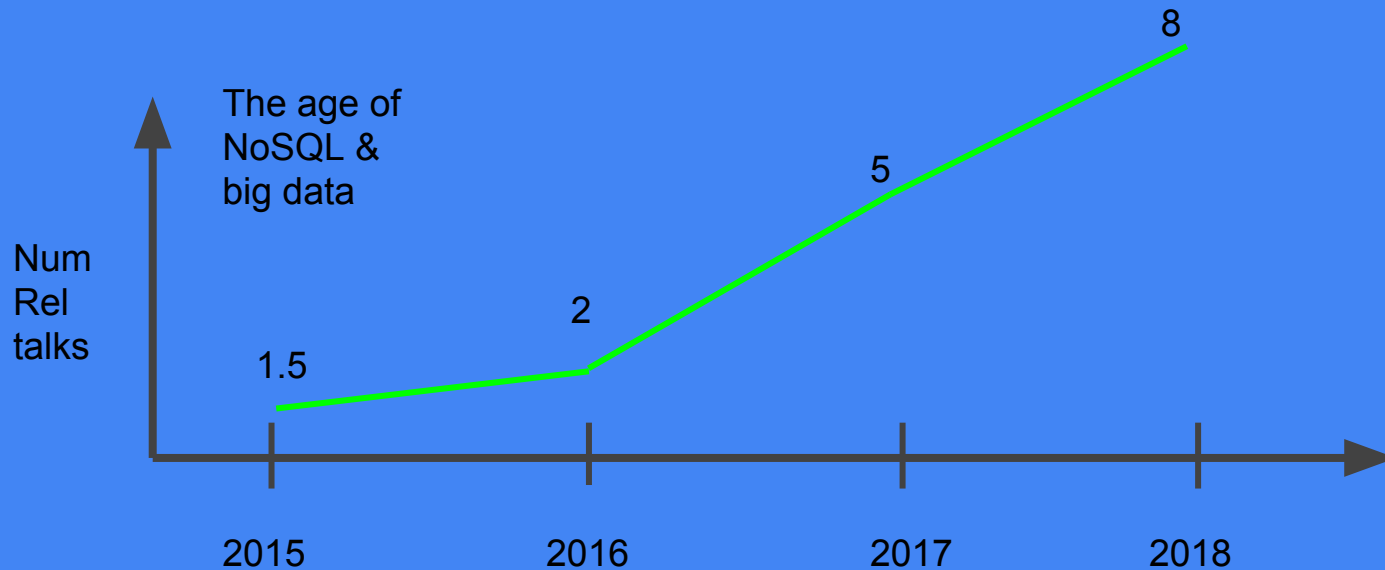


Elasticsearch LTR plugin

Bloomberg announces

Integrated into Solr

Relevance talks (bbuzz)



Bloomberg announces

Integrated into Solr

Elasticsearch LTR plugin

LTR brought the cool factor

Learning to Rank attracts people into search and relevance field

Incentive to

- Learn relevance best practices (judgment lists, etc)
- Advocate for relevance at work
- Go to cool conferences!



LTR plugin has good opinions

We take for granted many of the assumptions / opinions:

- Features are queries
- Search eng. store and log features
- Models you should know
(LambdaMART, rand. forest, linear, etc)
- Common language for practitioners
(not just academics)



Community finds LTR best practices

When getting LTR running is easier, the space is explored more, we all learn faster



Turns out... The best genetic algorithm is all of us trying to get LTR to work!

Solr workable for advanced teams

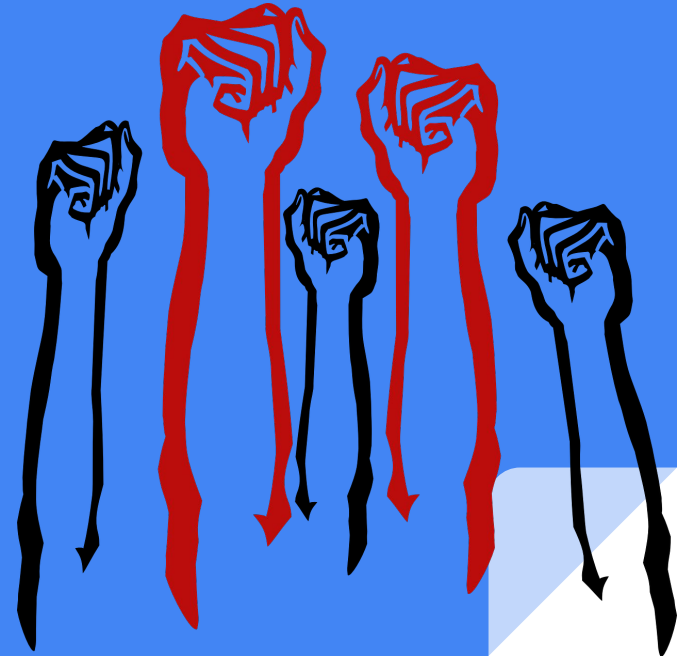


Solr (later ES) as a respectable system for advanced information retrieval

Teams can contemplate an LTR project

Not something out of reach for any search team

Functionality and knowhow within reach. And we can read the code!



Why did
Bloomberg do
this for us!?!


How much work did it take to build Bloomberg News LTR? (pre 2015)

- Conservative Guess:
 - 2 years,
 - 10 developers,
 - 150K salaries



€ 4 million!!

* No implied connection between Dr. Evil and Michael Bloomberg

How much was the investment to get LTR into Solr? (2015-2017)

- 102 comments
- OS plugin (2015)
- Jan 2016 - Apr 2017 (1.25 years)



€ 500k?

... evangelism time, travel... (2015 - today)

- Major conferences
- Active at meetups
- Demos (LTR for bees)
- Solr mailing list help
- Documentation
- JIRA...



€ 250-500k?

Why!?!?



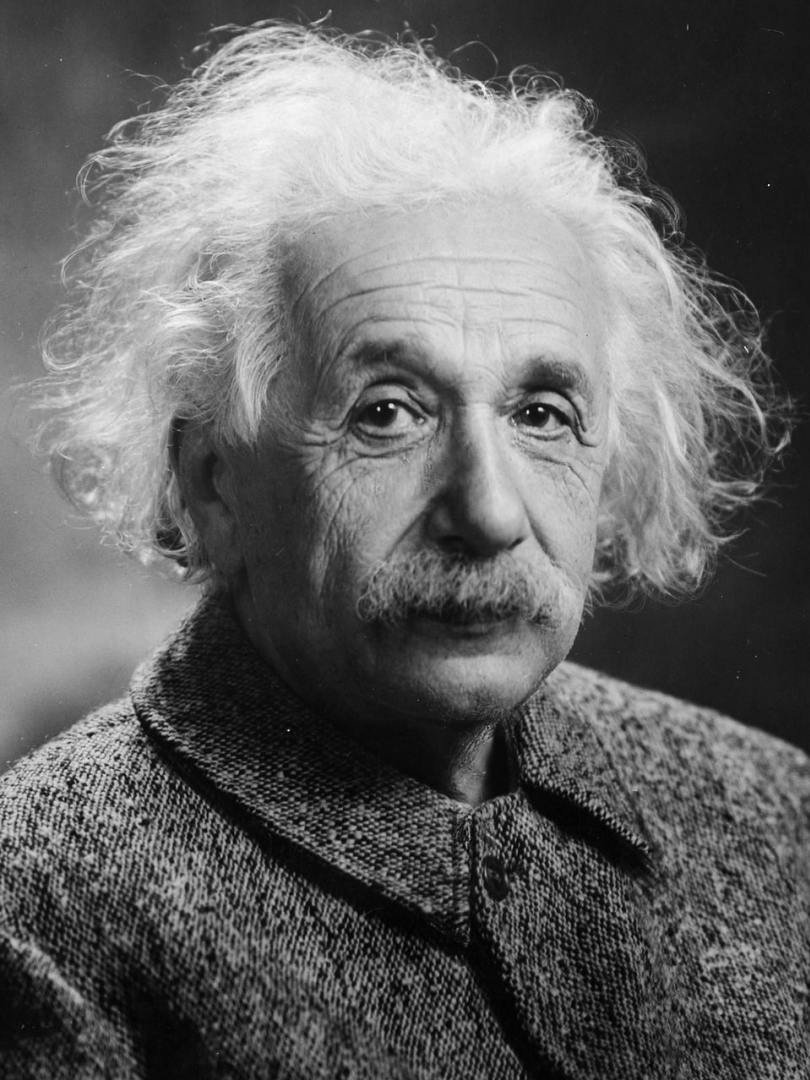
You: "We're going to spend €5 million dollar on functionality, and give it away"

Your boss's face



...Community Members...





Why was this a
genius move for
Bloomberg's
business?

Recruiting

Maintenance

open source contribution benefits

Collaboration

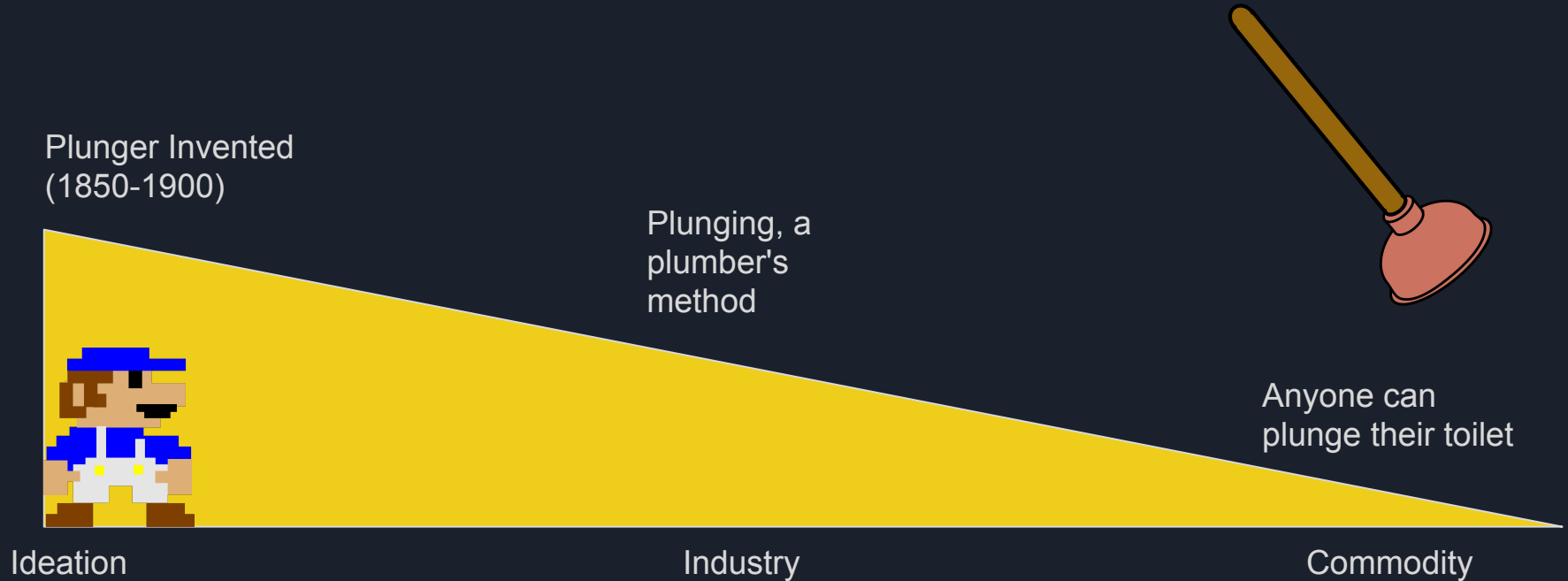
New features



...Also...



We need to commoditize repeat/busy work

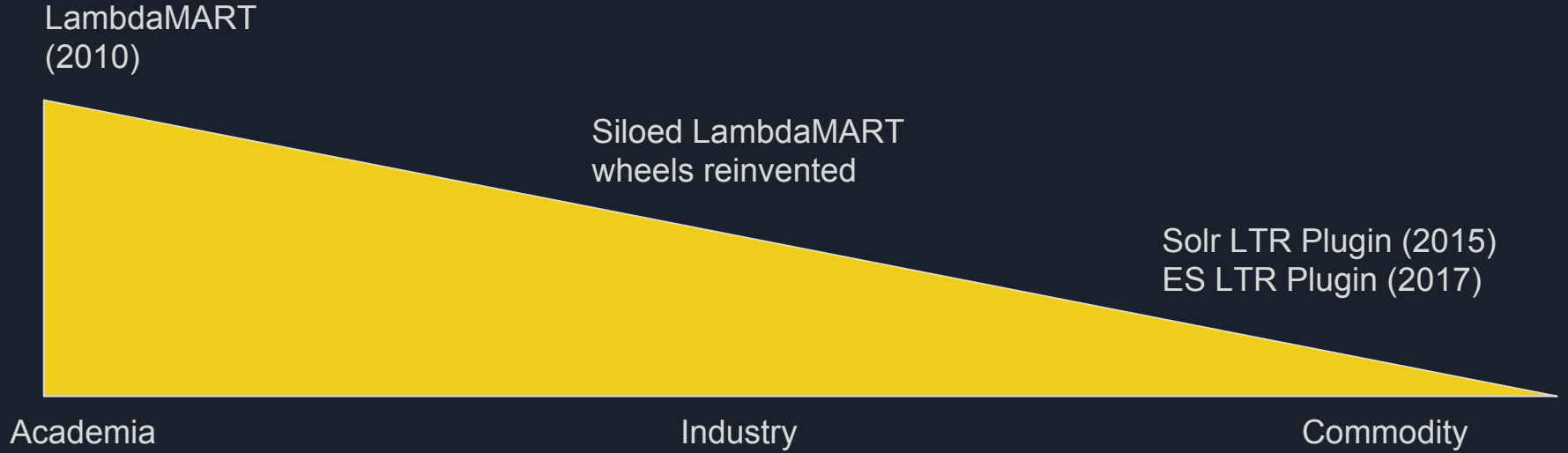


From original Haystack note:

https://docs.google.com/presentation/d/1wTZpwvTy9eYFKfPemBtblArZFSIGFiOo7QhEjkTX2a8/edit#slide=id.g374b6d60f3_0_85

#haystackconf

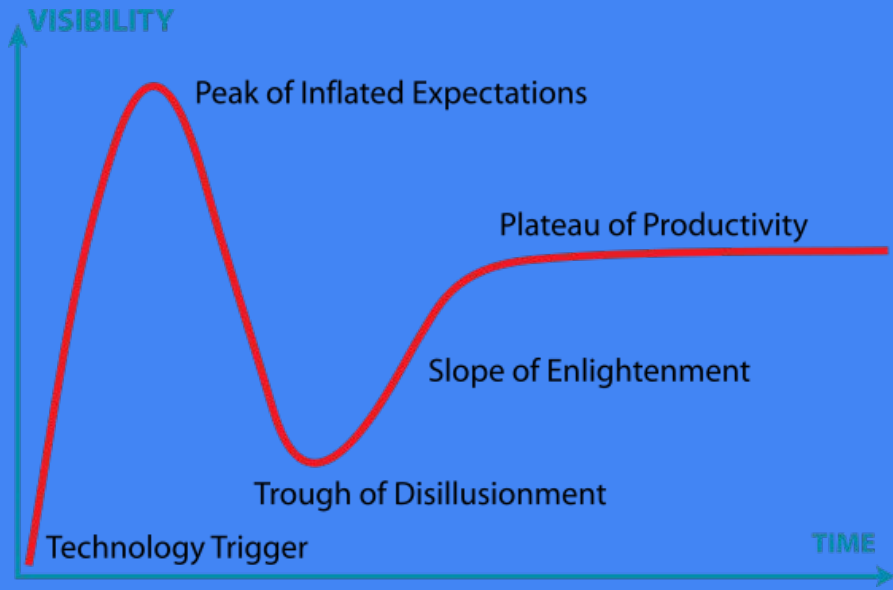
LTR building blocks getting easier



From original Haystack note:

https://docs.google.com/presentation/d/1wTZpwvTy9eYFKfPembtblArZFSIGFiOo7QhEjkTX2a8/edit#slide=id.g374b6d60f3_0_85

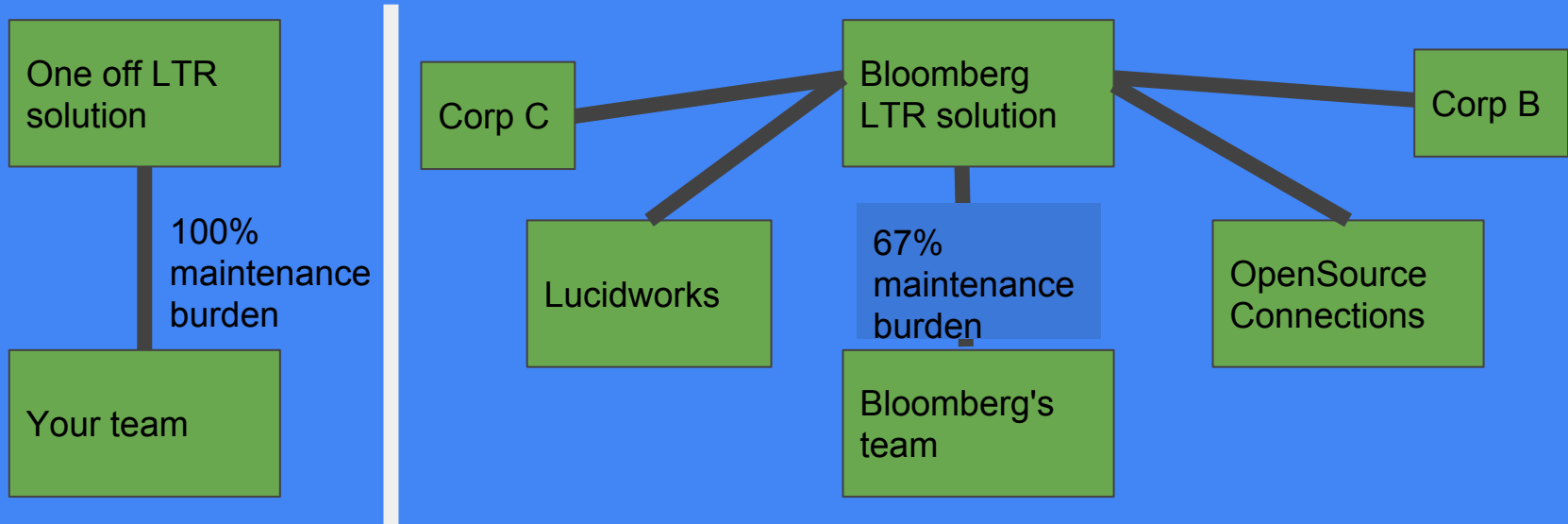
#haystackconf



Especially given
tech keeps
moving faster
and faster!

Where will your solution built
in 2014 be in 2019?

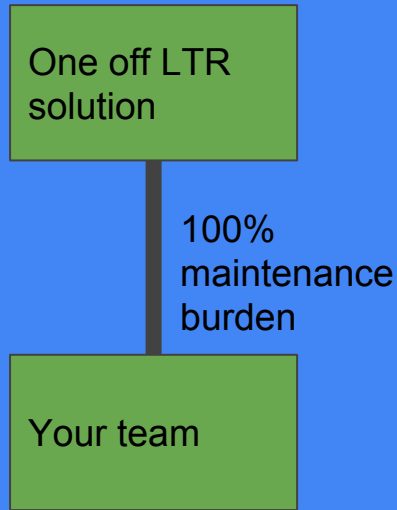
Dependency chain



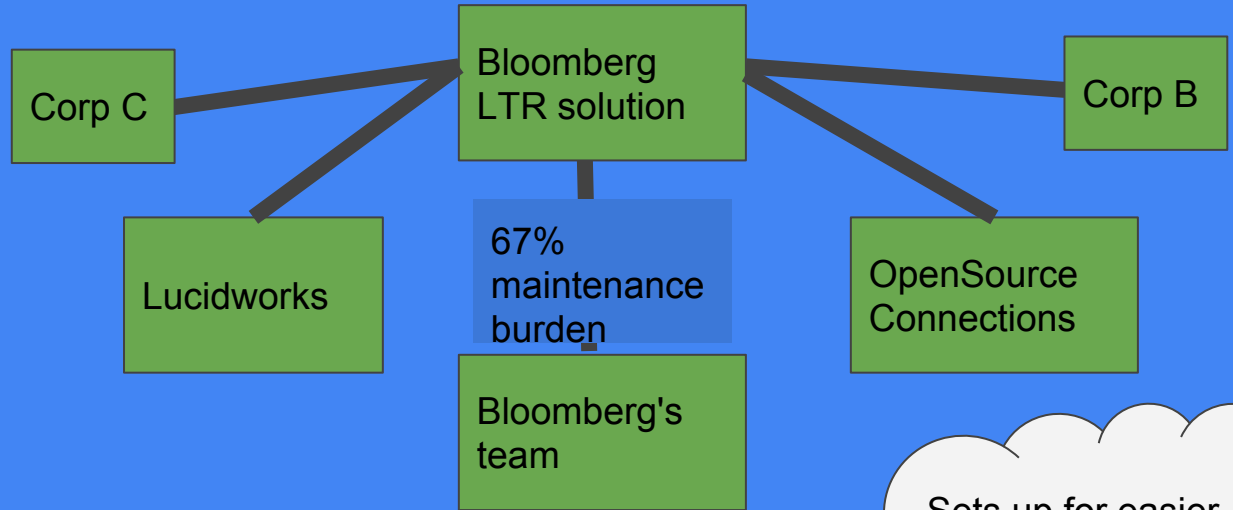
Pull of gravity: your team maintains everything

Pull of gravity: community maintains everything

Maximal reuse



Pull of gravity: solve one applications problems



Pull of gravity: Address broader range of use cases

Sets up for easier reuse **internally across Bloomberg**

Really this rule applies globally

One off LTR solution

100% maintenance burden

Your team

How good is your team's internal documentation?

Corp C

Bloomberg LTR solution

Corp B

Lucidworks

67% maintenance burden

Bloomberg's team

OpenSource Connections

How good is an open source projs documentation?

Sets up for easier reuse **internally across Bloomberg**

If you spent €5 mill 2014, which position would you rather be in in 2019?

One-off LTR solution

Almost obsolete

Your team stuck maintaining

Hard to apply to your co's other search probs

Bloomberg LTR solution

THE way of doing LTR

Community maintained: your team moves on to bigger challenges

Easier for others at your co. team to adopt

Bloomberg set the direction

We take for granted how LTR is supposed to work, because of Bloomberg's business decision



Probably part of a larger strategy



Berlin Buzzwords

@berlinbuzzwords

Following



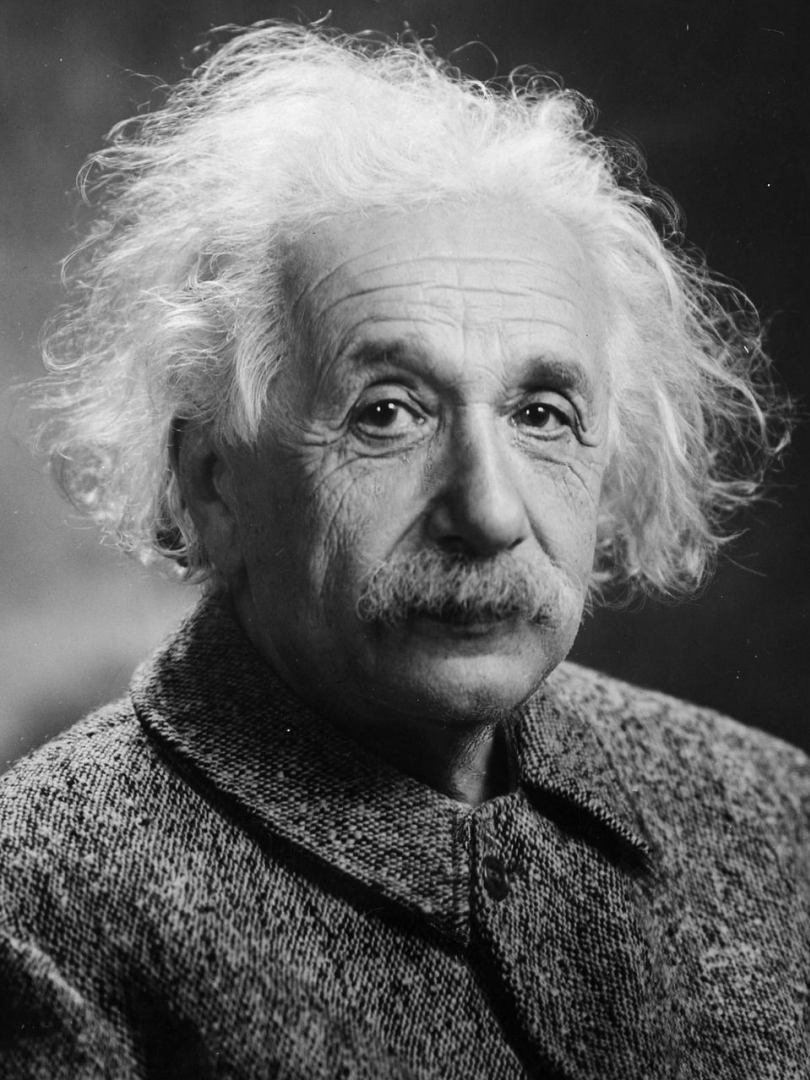
At #bbuzz 17 Ken LaPorte discussed how a vibrant search ecosystem was built at Bloomberg youtu.be/zag4IrC4z1k



Mid-size companies can't
maintain the internal search
ecosystem of a Google

But as a community we can





What's your
genius idea?

What initiative are you contemplating?



Integrate analytics

Use neural-NLP



Deal with
marketplace
dynamics

In 5-7 years, where would you rather be with this investment?

One-off
solution

Almost
obsolete

Your team stuck
maintaining

Hard to apply to
your co's other
search probs

Open Source
solution

THE way of
doing X

Community maintained:
your team moves on to
bigger challenges

Easier for others
at your co. team
to adopt



If we have seen farther than others...

It's because we've
stood on
shoulders of
giants

Sometimes we can be our own giants

