

# Open Source Search

- I'm Charlie Hull, co-founder and Managing Director of Flax
- We've been building open source search applications since 2001
- I'm going to tell you why and how you should use open source search engines



# So what's a search engine for?

- You'll know about Google but...
- ...what about searching your own data?
- Examples:
  - Product search for e-commerce
  - Article search for a newspaper website
  - Intranet search for a business
  - Search as part of a larger workflow system



# What about closed source?

- There used to be lots of (expensive) options
- ....but most of them were sold to big companies
  - Microsoft bought FAST
  - Oracle bought Endeca
  - Dassault Systemes bought Exalead
  - HP bought Autonomy
  - IBM bought Vivisimo



# So what's wrong with that?

- It's hard to buy *just search*
- Good support can be harder to find
- Prices have changed
- Can you be sure of the future?
- ...so many are moving to open source



# So what's open source?

- Source code is freely available for download
- Large community maintains the software
- Open development process
- Professional support is available



# What are the risks?

- It's insecure!
  - *All software has bugs, at least these are open*
- There's no-one in charge!
  - *No, development process is tightly controlled*
- It doesn't have the features I need!
  - *Development is rapid*
  - *You can build it yourself or engage others*
  - *Features & performance are comparable to closed source*
- There's no support!
  - *Plenty of support options available up to 24/7*



# So what are the options?

- Two options:

- Apache Solr

- Elasticsearch



elasticsearch.

- Both are applications with HTTP interfaces

- Both are based on Apache



- By the way, Twitter use Lucene to index **half a trillion tweets....**



Solr





# Solr - Pros



- Established community
- Development under Apache model
- Stable for very large scale deployments
- Lots of documentation, examples and books



# Solr - Cons



- Large and complex ecosystem
- Can be hard to get started with
- Many, many options (which you won't need)
- XML + Java + Zookeeper

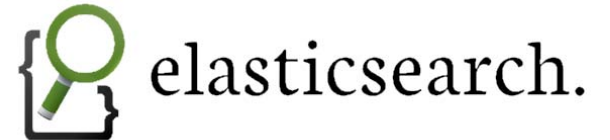




elasticsearch.



# Elasticsearch - Pros




- Highly active & growing community
- Easy to get started with
- ELK suite allows scalable log analysis
- Add-ons for security & alerts



# Elasticsearch - Cons



- Development managed by one company  elastic
- Known issues with data loss
- Much less documentation, examples
- Add-ons aren't always open source
- Java, JSON, Groovy



**But!**



- They're both very solid platforms for search
- We've used both successfully for client projects
- Performance and features broadly similar:

<http://www.slideshare.net/charliejuggler/lucene-solrlondonug-meetup28nov2014-solr-es-performance>



# So how do I get started?

- Get some books and/or training
- Download the software for free
- Build an example project
- Ask questions of the community
- Easy, eh?



# Actually it's not that simple!

- *All* search projects need proper planning
- User needs should drive the overall process
- You may need other software to condition data
  - e.g. Apache Tika to extract text from documents in many formats
  - e.g. Nutch to crawl websites
- Consider a PoC as a first step
  - Small, representative sample of data
  - Basic search UI
  - Get lots of feedback





# Who is available to help?

- Search consulting companies like Flax
- <https://wiki.apache.org/solr/Support> (Solr)
- Lucidworks Fusion, based on Solr
- Elastic.co offer support for Elasticsearch



# Use the community

- Mailing lists for Solr and Elasticsearch
- Search Meetups - [www.meetup.com](http://www.meetup.com)
- Conferences:
  - Lucene Revolution, Austin, Texas, 13<sup>th</sup>-16<sup>th</sup> October
  - Elasticon (US was 9<sup>th</sup>-11<sup>th</sup> March, European one coming?)
  - Berlin Buzzwords
  - Apachecon



# Try to give back

- Contributing code as open source
- Not compulsory!
- Solr – via Apache JIRA
  - <https://issues.apache.org/jira/browse/LUCENE>
  - <https://issues.apache.org/jira/browse/SOLR>
- Elasticsearch – Github
  - <https://github.com/elastic/elasticsearch>
  - ....*but Elastic control any patches*



# Thankyou

- Open source is powerful, freely available and gives you back control
- There's plenty of help available if you need it
- Open source model leads to rapid development and bugfixing
- Ask me anything!
  - [charlie@flax.co.uk](mailto:charlie@flax.co.uk)
  - [@FlaxSearch](https://twitter.com/FlaxSearch)

