

Search Indexer Task Force - First meeting

Regular Attendees

- Jim Blake
- Don Elsborg
- John Ferreira
- Huda Khan
- Ted Lawless [blocked URL](#)
- Jing Wang
- Stephan Zednik
- ([blocked URL](#)) Indicates who will be taking minutes

General

1. Call-in details: Dial-In Number: , Participant code: **TBD**
- 2.

Agenda Items

		Lead
1	What are we doing here? Review the charge, individual goals, discussion	Jim
2	How the indexer works now	Jim
3	What happened at the Hackathon	Don, Huda
4	What do we want for the future?	All
5	Agenda for next meeting	All

Action Items from this meeting

- List action items here and responsible person

Minutes

Jim will go home if it's the Jim Blake Show.

- What are we doing here?
 - This group is convened with a "Task Force Charge".
 - The group was formed after discussions between Jim and Don about search indexing performance and results. Jim decided to form a task force around these discussions.
 - Jim's initial goal - now that the search indexer is more configurable we should talk about it and document it to assist others with configuring it.
 - Don is interested in looking forward and wants to look at the work on facets done at the hackathon (October 2014 in Cornell).
 - John mentions that some sites are interested in Blacklight or Drupal as a VIVO UI. Also VIVO searchlight could use the VIVO search index.
 - Jim asks about Blacklight - Huda is working on using Blacklight to access a VIVO managed Solr index. John says if patterns are defined (field names, facets) it would be easier for other systems to use the VIVO Solr index.
 - Huda says that documenting the schema or how to change it is one thing but then populating the custom schema is another.
- What prompted VIVO to use Solr?
 - Don would like to know what the problem was that prompted the use of SOLR in VIVO. If it's because of slow aggregations in the triple store - is this a jena specific problem? dbpedia returns counts of millions of objects in short time frames.
 - Initially Lucene was used
 - Original reason was to facilitate free text querying
 - Then used search index to help in performance areas (e.g. how many faculty members total) where the triple store isn't as fast as search index
 - Now used for cache control as well.
 - At some point the change was made from Lucene to Solr, we think to support faceting.
 - Don asks if Solr is still relevant if a faster triple store backed VIVO.
 - Jim says that forms use autocomplete via Solr throughout.
 - John says Solr is used for discovery and search rather than storing all of the content.
 - Stephan says ranking of free text query comes from Solr. Standard triple stores don't do that well.
 - Example - how do I rank certain types (People) higher than others (Research Areas). Documentation needed.
- How the indexer works now

- John - do we want a UI to help configure Solr? Ted mentions multiple backend options (e.g. Elasticsearch) is now possible so that makes it difficult. Don prefers config in files that can be automated. Jim mentions that config is now N3 files so there is potential to edit that within Vitro/VIVO. Jing thinks this would be a next step/out-of-scope - document use case is in scope.
- Task force members - why are you here and what do you want out of it?
 - Don - direction and understanding of how things are configured now and why. Long term want to have faceted search and browsing. Interested in more flexibility and functionality.
 - Stephan - interested in faceted browsing for RPI VIVO. Wants to share specific use cases with community rather than implementing it himself and trying to push into core. Cares about administration and making facets configurable, so you don't need a Java programmer experienced with VIVO to make changes. Wants to document other use cases.
 - Huda - working with VIVO's search index in a couple of projects. Wants to be part of developing a solution and learning more.
 - John - interested in documenting and sharing how the search index can be customized. Specific use case - AgriVIVO uses Drupal and having an understandable or solid Solr index is important.
 - Jing - interested in use case regarding discovery research and scholarship beyond publications. Interested in how ontology might be helpful in facilitating faceting and more complicated search. Wants to understand how the current VIVO index works.
 - Ted - would like to understand more how it works so he can tweak solr to configure facets. interested in blacklight. Interested in using VIVO/vitro as a tool to store data that is searchable.
- What happened at the Hackathon
- iFest
 - John asks if this task force will be represented or do work at the iFest
 - Stephan asks if it would be a good place to document use cases
 - Jing suggests doing some preliminary work on use cases before iFest and using face-to-face time to document those.
 - Don volunteers to start a document for use cases and will circulate to mailing lists. Will try to get something out today.
 - in particular -- we should also have a use case doc for our task force
 - Jing suggests that each of us contribute one use case. Stephan says he will add his.
 - John sees iFest as an opportunity to get more use cases from the community.
- What do we want for the future?
 - Stephan thinks we should collect use cases from the community. don seconds
 - melbournes "find an expert" by Simon Porter
 - auto-complete in form (Ted, Brown)
 - vivo searchlight
 - discovery of research resources
 - how to configure the search configuration
 - Don - How does our search index work dovetail with work done by other projects associated with VIVO - eg blacklight. How can we have more efficient dovetailing across projects in the future
- Task force outcome / goals / deliverables
 - Jim says focus on attainable goals and what a follow-up group may be able to accomplish.
- Agenda for next meeting
 - Documenting community use cases
 - Documenting how existing search indexing system works
 - Clarify deliverables and how do we produce them
 - Jim will send doodle poll for next meeting time
 - Don - how do we know what is going on in related projects like Blacklight or Funnelback?
 - John suggests stepping back and thinking about faceting in general and not targeting end user applications like Drupal or Blacklight.