

iLab

IR Experiment Environment

Feb 2009, Jinyoung Kim

Introduction

- What is it?
 - iLab is a ruby-based framework (API, executables) in which IR researchers can design, perform experiment.
 - iLab interacts with search engines to run retrieval, collect and analyze results, generate report.
- Why Ruby?
 - It's a dynamic scripting language in which OO and functional paradigms are well-combined.
 - Plus, it's easy to learn and fun to use.

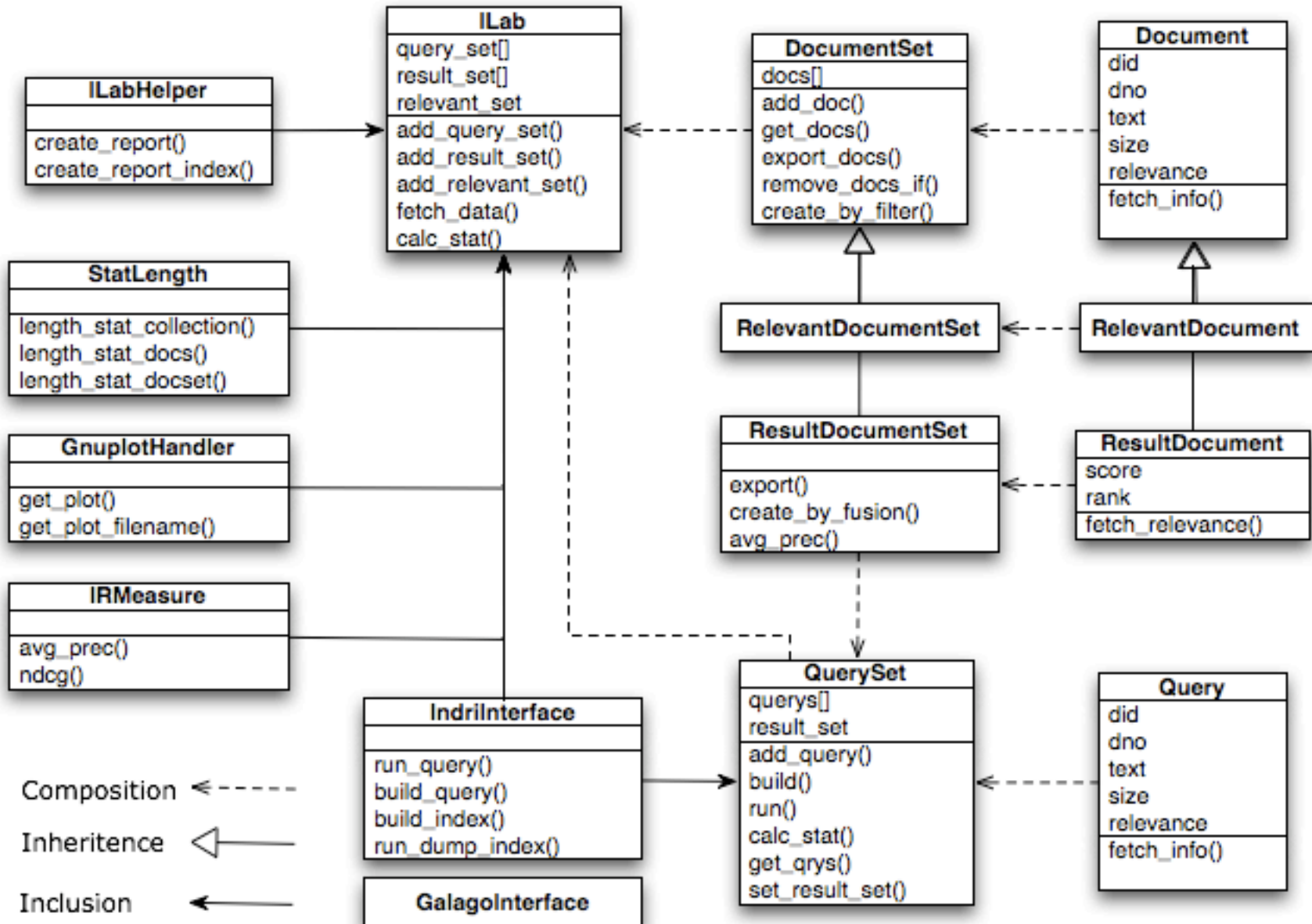
Motivation

- IR research involves lots of repeated work
 - e.g. Query generation, retrieval, data analysis
 - Scripting helps, yet not in a graceful way
- More structured approach will enable:
 - Automatic, repeatable ,scalable IR experiment
 - Knowledge transfer between researchers

Features

- OO-Wrappers for major elements of IR research
 - Document (set), Query (set)
- Routines for IR Experiment Procedure
 - Query generation (template-based)
 - Perform retrieval (local / remote - cluster)
 - Get statistics (most IR metrics are embedded)
 - Generate report (template-based, Gnuplot support)
 - Perform significance test (interface to R)
- Generic interface to search engines
 - Extendable to support more than Indri

Class Architecture



Designing an Experiment

Initial Setting

- Global Path

Choose Collection

- Collection Path, Topic, Qrel,

Choose Retrieval Method

- Type(title/desc/narr), Template(QL/DM), Smoothing, Prior, ...

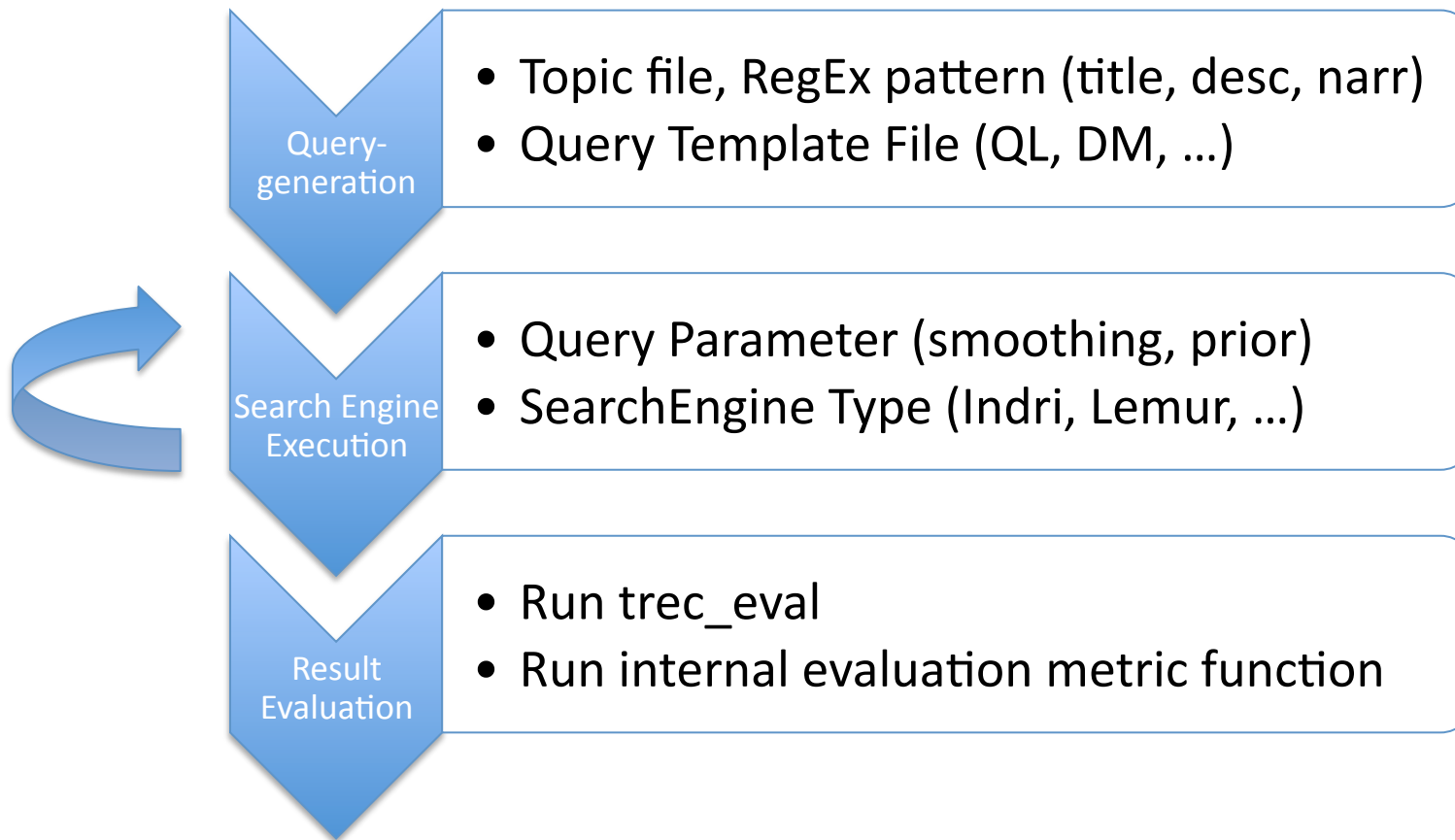
Choose Experiment

- Length Analysis, Query-wise Analysis

Define Report & Run

- Numbers, Chart, Graph, ...

Running Experiment



Ad-hoc Script vs. iLab

Code is hardly reused

- Everything in iLab is object, parameterized, template-based.

They don't scale well

- iLab experiments can be combined to build a more complex one.

They're fragile

- iLab is based on Test-Driven Development.

They're hard to debug

- Ruby provides interactive debugging shell.
- iLab logs everything for you.

You have to put efforts

- Everything is one-click away.

How To Get & Use

- SVN Repository
 - `svn+ssh://`
- RDoc Source Document
 - <http://www.lifidea.com/>
- Start-up & User Guide
 - <http://www.lifidea.com/>

Demo

Future Work

- Parallelism using cluster
- More search engine interfaces
 - Galago, Lucene, and so on