I love Lucy
and Dezi too

Apache Lucy
Dezi
DBIC
101

By Logan Bell

Goals of Talk

- Give some background

- Create an Index

- Setup a search service

- Query the service

Terms

- Lucy - An indexer written in C, loosely ported from Lucene

- Dezi - A client server package that queries Lucy. Uses Swish3 for parsing when indexing.

- Swish3 - Swishe compatible parser. Will parse incoming pdf, docs, html, xml, and others.

- DBIC - Popular ORM for Perl

Lucy Background

- An indexer written in C, loosely ported from Lucene

  - Fast full text search

  - Results based on relevance

  - Proximity searches

  - Eventually be used in any dynamic language

State of Lucy

- Still in apache incubation

- Active development

- Perl is the only bound language at this time.

- Ruby is in development.

Stand alone Lucy Indexer

```perl
#! /usr/bin/env perl
# Example from https://metacpan.org/module/Lucy::Simple

use Lucy::Simple;

my $index = Lucy::Simple->new(
    path     => './index',
    language => 'en',
);

my %source_docs = (
    'Foo' => 'I am a page about birds that fly high in the sky',
    'Bar' => 'This is a page about whales that hold their breath in the sea'
);

while ( my ( $title, $content ) = each %source_docs ) {
    $index->add_doc({
        page     => $title,
        content  => $content,
    });
}
```

Stand Alone Lucy Query

```perl
#! /usr/bin/env perl

use strict;
use warnings;

use Lucy::Simple;

my $index = Lucy::Simple->new(
    path     => './index',
    language => 'en',
);

if ( my $query_string = $ARGV[0] ) {

    my $total_hits = $index->search(
        query      => $query_string,
        offset     => 0,
        num_wanted => 10,
    );

    print "Total hits: $total_hits\n";
    while ( my $hit = $index->next ) {
        print "$hit->{page}\n",
    }
}
```

Dezi

- A client server package that queries Lucy. Uses Swish3 for parsing when indexing and OpenSearch for return result sets.

  - Plack and restful JSON api.

  - Parses xml, binary, html, pdf, doc.

  - Written by Peter Karman, maintainer of SWISH3.

Installation?

- Extremely easy to install

  - Lucy: cpanm Lucy

  - Swish3:

    - svn checkout http://svn.swish-e.org/libswish3/trunk libswish3

  - Dezi: cpanm Dezi

Test it Out

% dezi &

% echo '<doc><title>bar</title>hello world</doc>' > test.xml

 % dezi-client test.xml

 % dezi-client -q bar

Swish3

- Based on an older parser/indexer project SWISHE.

- Parses doc, pdf, html, xml, and other formats.

- Supports filters for any transitions.

  - Parse an image, transform it into xml. Examples:

    - SWISH::Filters::ImageToXml

    - SWISH::Filters::ImageTypesToXml

    - SWISH::Filters::ImageToMD5Xml

DBIC

- A popular perl ORM package.

- DBIx::Class::Indexer::WebService::Dezi

- DBIx::Class::IndexSearch::Dezi

Example Schema

```
__PACKAGE__->load_components(qw[
    Indexed
    IndexSearch::Dezi
    PK::Auto
    Core
    TimeStamp
]);
```
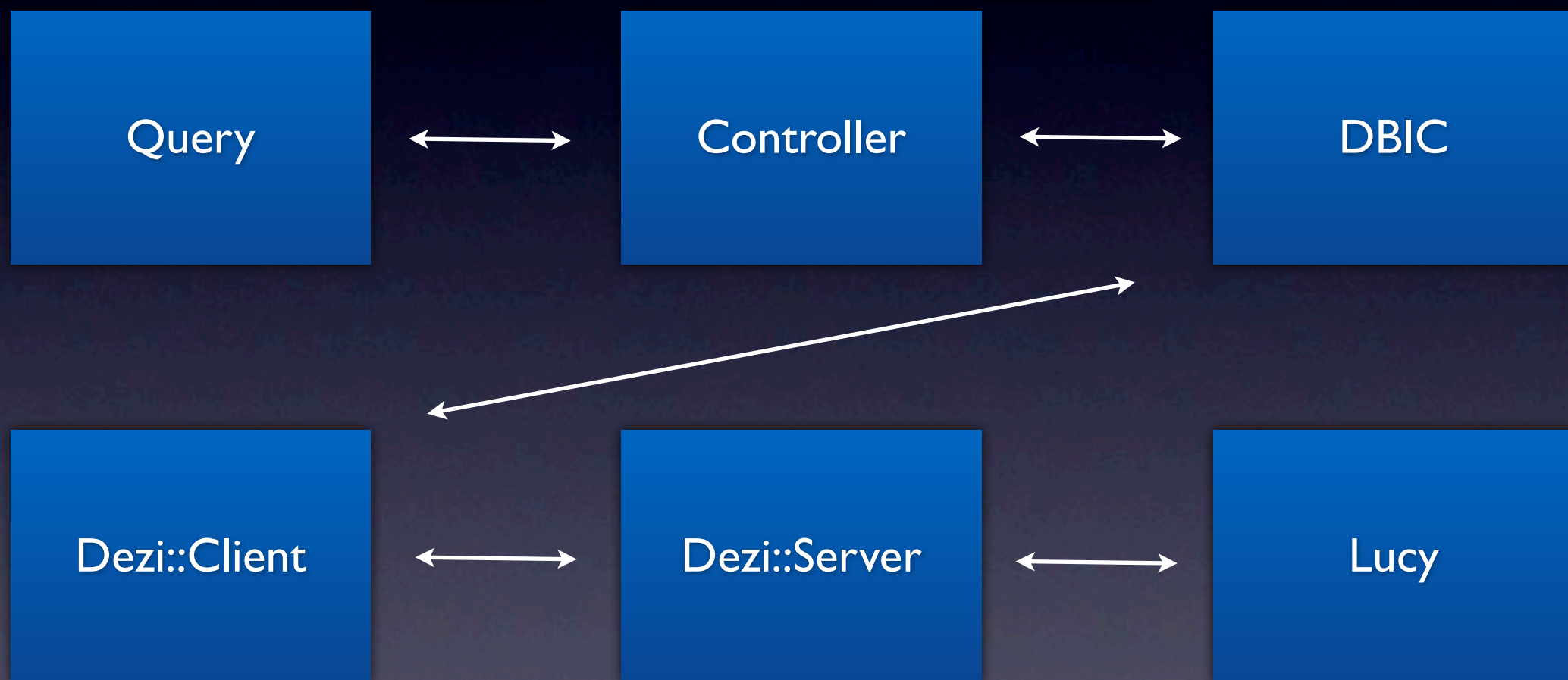
Example Schema

```
# Set Indexer
__PACKAGE__->set_indexer('WebService::Dezi', { server => 'http://localhost:6000',
content_type => 'application/json' });
__PACKAGE__->table('data');
__PACKAGE__->add_columns(
    id => {
        data_type       => 'id',
        is_nullable     => 0,
    },
    key => {
        data_type       => 'varchar',
        is_nullable     => 0,
        indexed         => 1
    },
);
__PACKAGE__->set_primary_key('id');
__PACKAGE__->resultset_class('DBIx::Class::IndexSearch::ResultSet::Dezi');
__PACKAGE__->belongs_to_index('Dezi::Client', { server => 'http://localhost:6000', map_to
=> 'key');
}
1;
```

Freestock and Lucy

- Implemented various flavors of SWISH filters

  - SWISH::Filters::ImageToXml

  - SWISH::Filters::ImageToMD5Xml

  - SWISH::Filters::ImageTypesToXml

- Created custom DBIC components to index and search against Lucy via Dezi

  - DBIx::Class::Indexer::WebService::Dezi

  - DBIx::Class::Indexer::IndexSearch::Dezi

Query Workflow

More Advanced Example

- https://github.com/logie17/LucyDeziTalk

- plackup -Ilib bin/run.pl

- dezi --dezi-config ./config/dezi.pl -p 6000

- http://localhost:5000

Further Reading

- http://incubator.apache.org/lucy/

- https://metacpan.org/module/Lucy::Docs::Tutorial

- https://metacpan.org/module/Dezi::Tutorial

- #lucy_dev (freenode)

Thanks

- #lucy_dev

- Peter Karman (author of DBIC)

- Marvin Humphrey (Founding PPMC of Lucy)