

Introduction to Galago Query Language

Introduction

- A document is viewed as a sequence of text that may contain arbitrary tags.
- A single context is generated for each unique tag name.
- An extent is a sequence of text that appears within a single begin/end tag pair of the same type as the context.
- Some contexts can be "title", "body" "h1" etc.

```
<html>
<head>
<title>Department Descriptions</title>
</head>
<body>
The following list describes ...
<h1>Agriculture</h1> ...
<h1>Chemistry</h1> ...
<h1>Computer Science</h1> ...
<h1>Electrical Engineering</h1> ...
</body>
</html>
```

body context:

```
<body> The following list describes ...
<h1>Agriculture</h1> ...
<h1>Chemistry</h1> ...
<h1>Computer Science</h1> ...
<h1>Electrical Engineering</h1> ...
</body>
```

h1 context:

```
<h1>Agriculture</h1>
<h1>Chemistry</h1> ...
```

Operators

- Ordered operator: #od:N(..)
- #od:N(..) terms must appear ordered with atmost N-1 terms between them.
- #od() : unlimited ordered window
- Try:

#od:1(iron maiden)

#od:1(new york)

#od:1(immanuel kant)

Operators (cont.)

- e.g. dog.title will look for “dog” in the title.
- #syn() : Create synonyms from two terms.
- e.g. #syn(#od:1(united states) #od:1(united states of america))

- Try

#syn(dog canine)

#syn(robert bob)

- Suggest more?

Belief Operators

- `#combine(..)`, `#weight(..)`.
- `#filter(..)`: this operator is similar to `#combine`, but with the difference that document must contain at least one instance of all terms.

Some examples

- `#combine(#syn(dog canine) training)`
- `#filter(aquarium #combine(tropical fish))`: filter documents that contain “aquarium” then rank them according to `#combine(aquarium #combine(tropical fish))`
- `#combine(#od:1(civil war) #uw:4(war independence))`
- `#weight(1.0 godfather 0.7 #od:1(don corleone))`