

Metaphor Detection

through

Term Relevance

Marc Schulder

Saarland University



Eduard Hovy

Carnegie Mellon University



Metaphor Detection

“And yet we stand together as we did two centuries ago.”

Metaphor Detection

“And yet we stand together as we did two centuries ago.”

Metaphor Detection Challenges

Required Knowledge

Metaphor Detection Challenges

Required Knowledge

Conceptual Mappings

Target Domain × Mappings × Source Domains

Metaphor Detection Challenges

Required Knowledge

Conceptual Mappings

Target Domain × Mappings × Source Domains

Selectional Preference Violation

Preferences × Argument Domains

Metaphor Detection Challenges

Required Knowledge

Conceptual Mappings

Target Domain × Mappings × Source Domains

Selectional Preference Violation

Preferences × Argument Domains

Typical Restrictions

Low Coverage

POS Limitations

Metaphor Detection Challenges

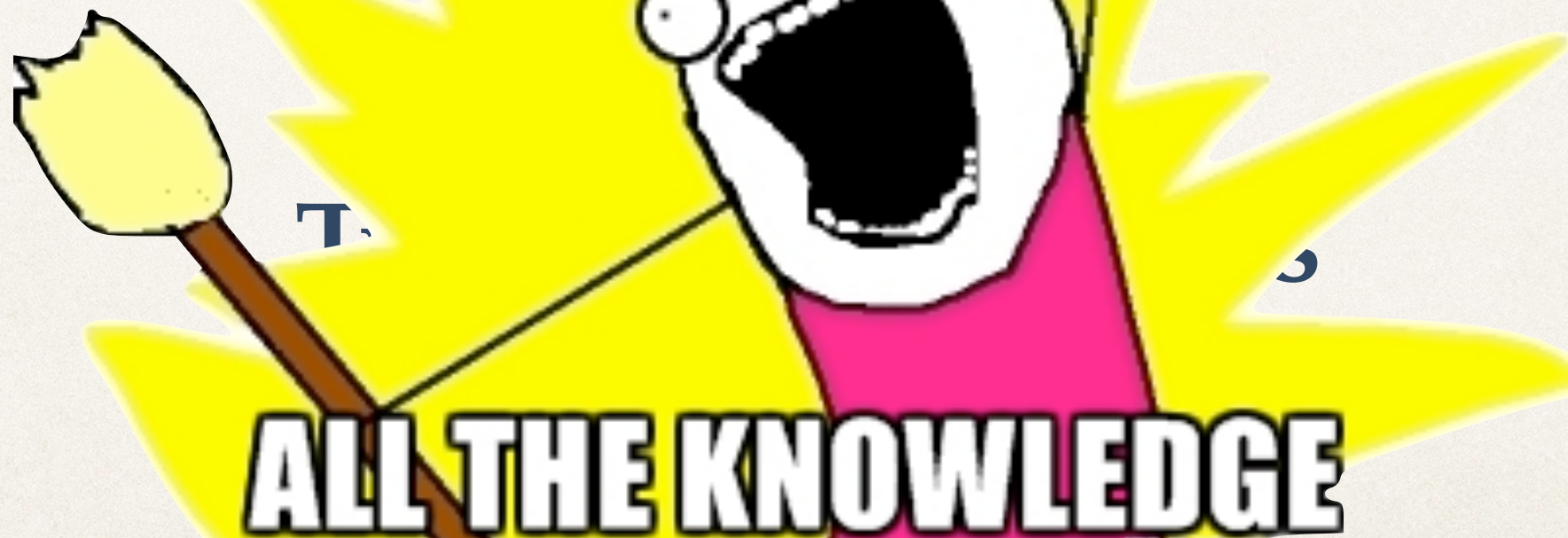
Required Knowledge

REQUIRE

Target Domain Source Domain

Selective Knowledge Information

Reference Domains



Metaphor Detection

Term Relevance

Metaphor Detection

Term Relevance

Simple

Metaphor Detection

Term Relevance

Simple

Robust

Metaphor Detection

Term Relevance

Simple

Robust

POS independent

Metaphor Detection

Term Relevance

Simple

Robust

POS independent

Target Domain only

Term Relevance

Hypothesis

If a word does
not fit in the context,
then it is probably
not meant literally.

Term Relevance Overview

Relevance Metric

Domain Data
(Web Corpus)

Evaluation
(Metaphor Corpus)

Basic Classifier

Multi-Feature Classifier

Term Relevance Metric

Is word common
in all domains?

yes

literal

no

Is word typical
for this domain?

yes

literal

no

metaphor

Term Relevance

TF-IDF

Term Frequency

How often does term appear in a document

Document Frequency

In how many documents does term appear

TF-IDF

Impact of term on document

term frequency \times inverse document frequency

Term Relevance

TF-IDF

Term Frequency

How often does term appear in a document

Document Frequency

In how many documents does term appear

TF-IDF

Impact of term on document

term frequency \times inverse document frequency

Term Relevance

TF-IDF

Term Frequency

How often does term appear in a document

Document Frequency

In how many documents does term appear

Domain Relevance

Impact of term on document

term frequency \times inverse document frequency

Term Relevance Metric

Is word common
in all domains?

yes

literal

no

Is word typical
for this domain?

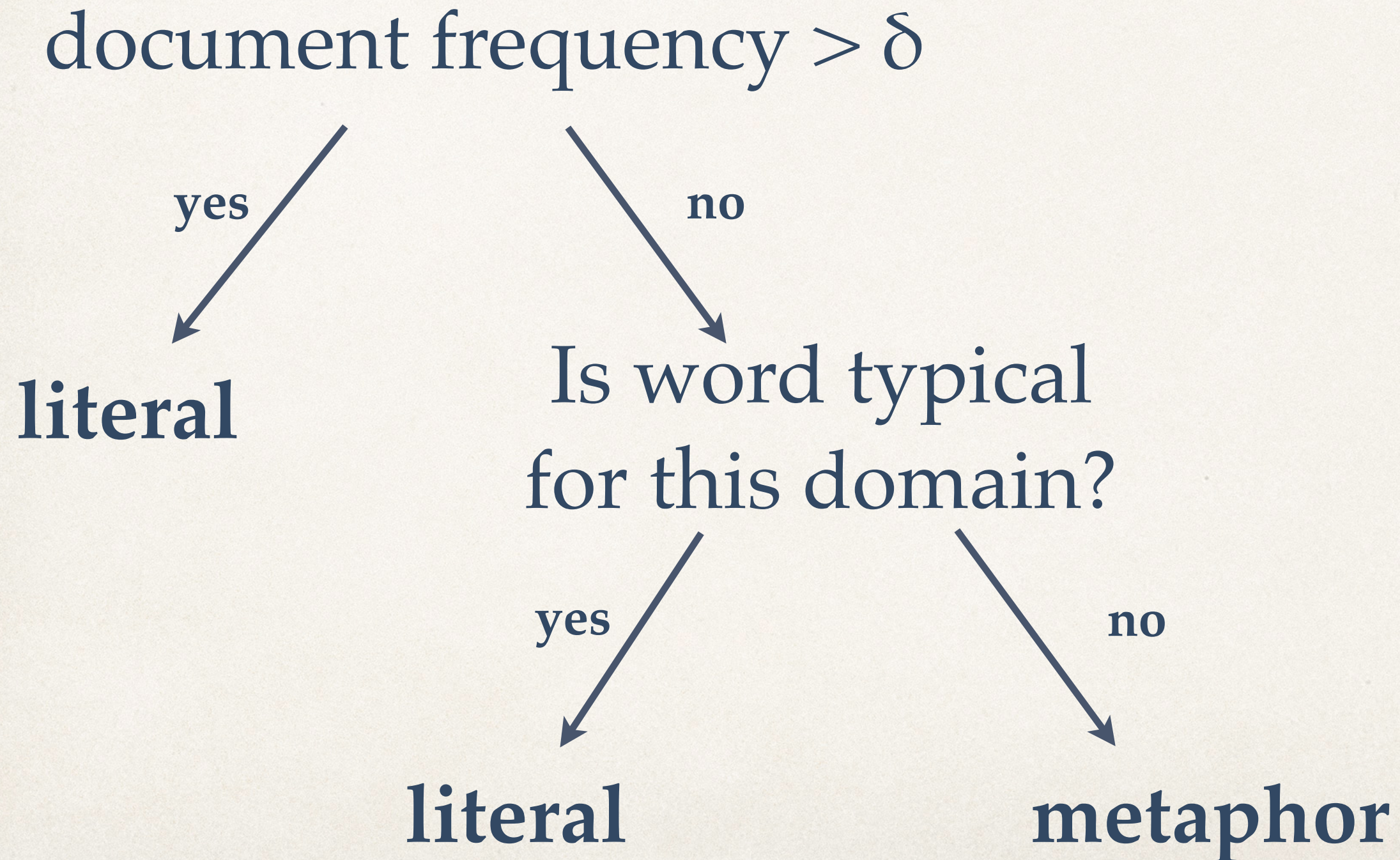
yes

literal

no

metaphor

Term Relevance Metric



Term Relevance Metric

document frequency $> \delta$

yes

no

literal

domain relevance $> \gamma$

yes

no

literal

metaphor

Overview

Relevance Metric

Domain Data

Evaluation

Basic Classifier

Multi-Feature Classifier

Domain Data

Web Corpus

ClueWeb-09

1 Billion Web Documents

500 Million English Web Documents

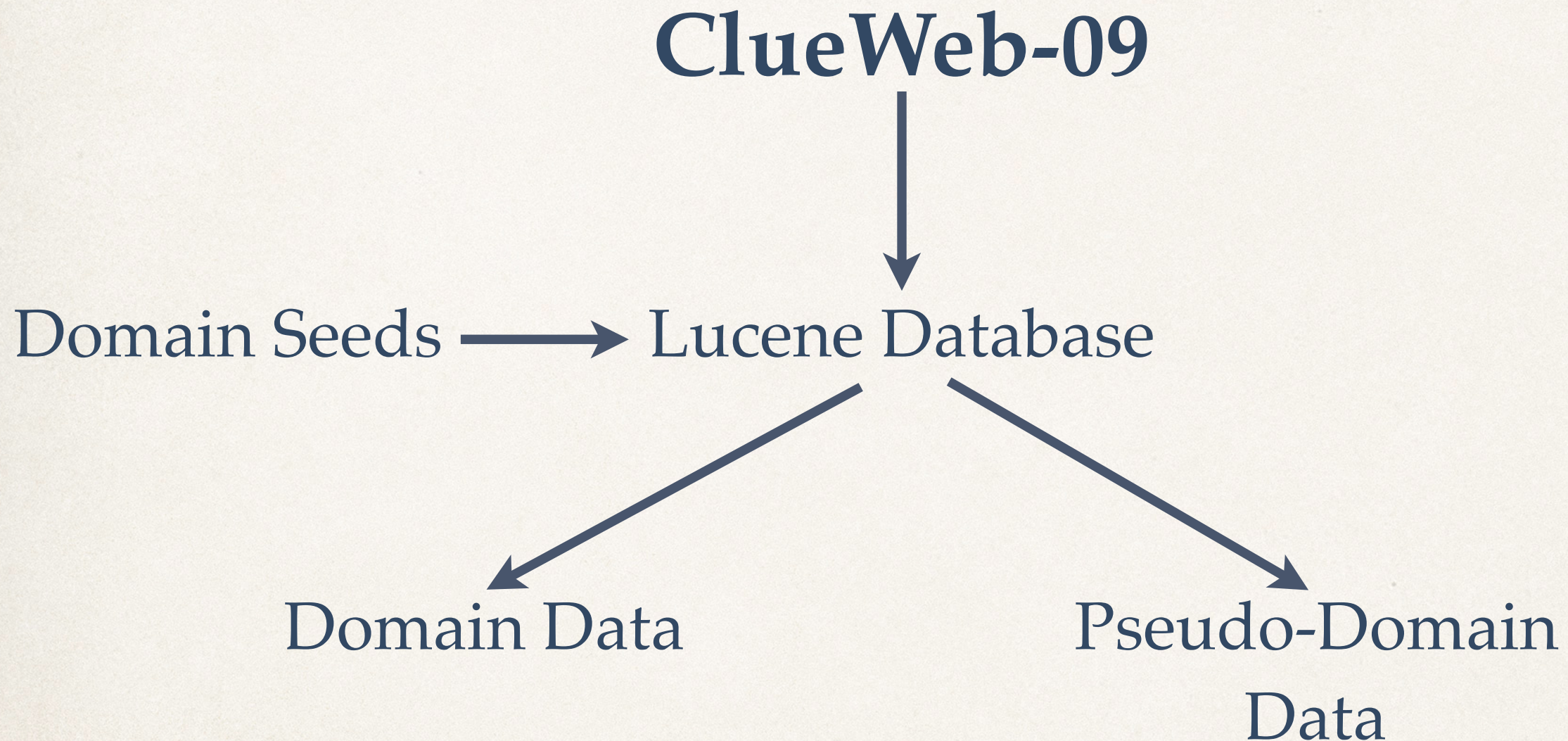
Segment en0000

3 Million Documents

1.8 Million Documents without Spam

Domain Data

Domain Clustering



Domain Data

Domain Clustering

ClueWeb-09

Legislative

pass law regulate
debate parliament

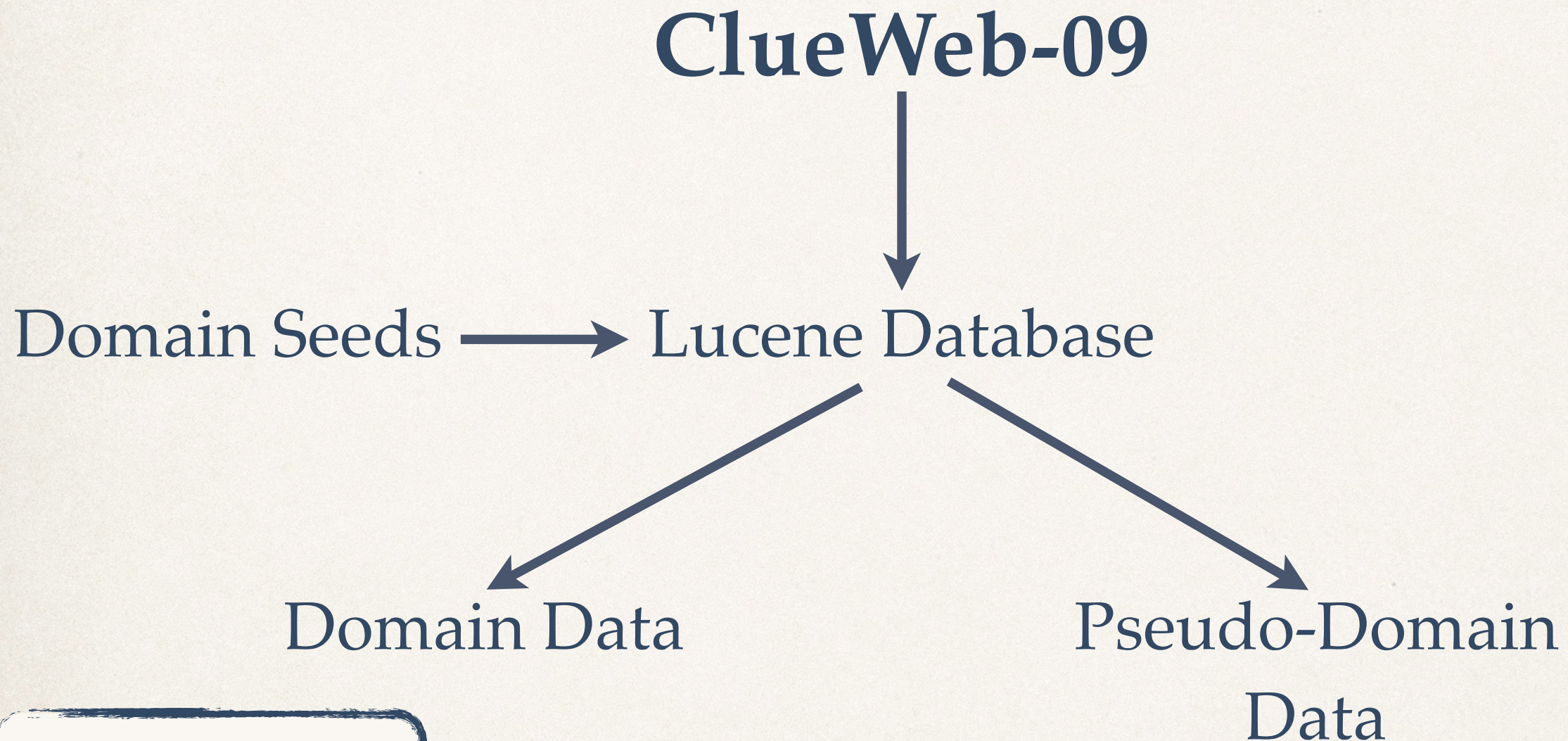
Lucene Database

Domain Data

Pseudo-Domain
Data

Domain Data

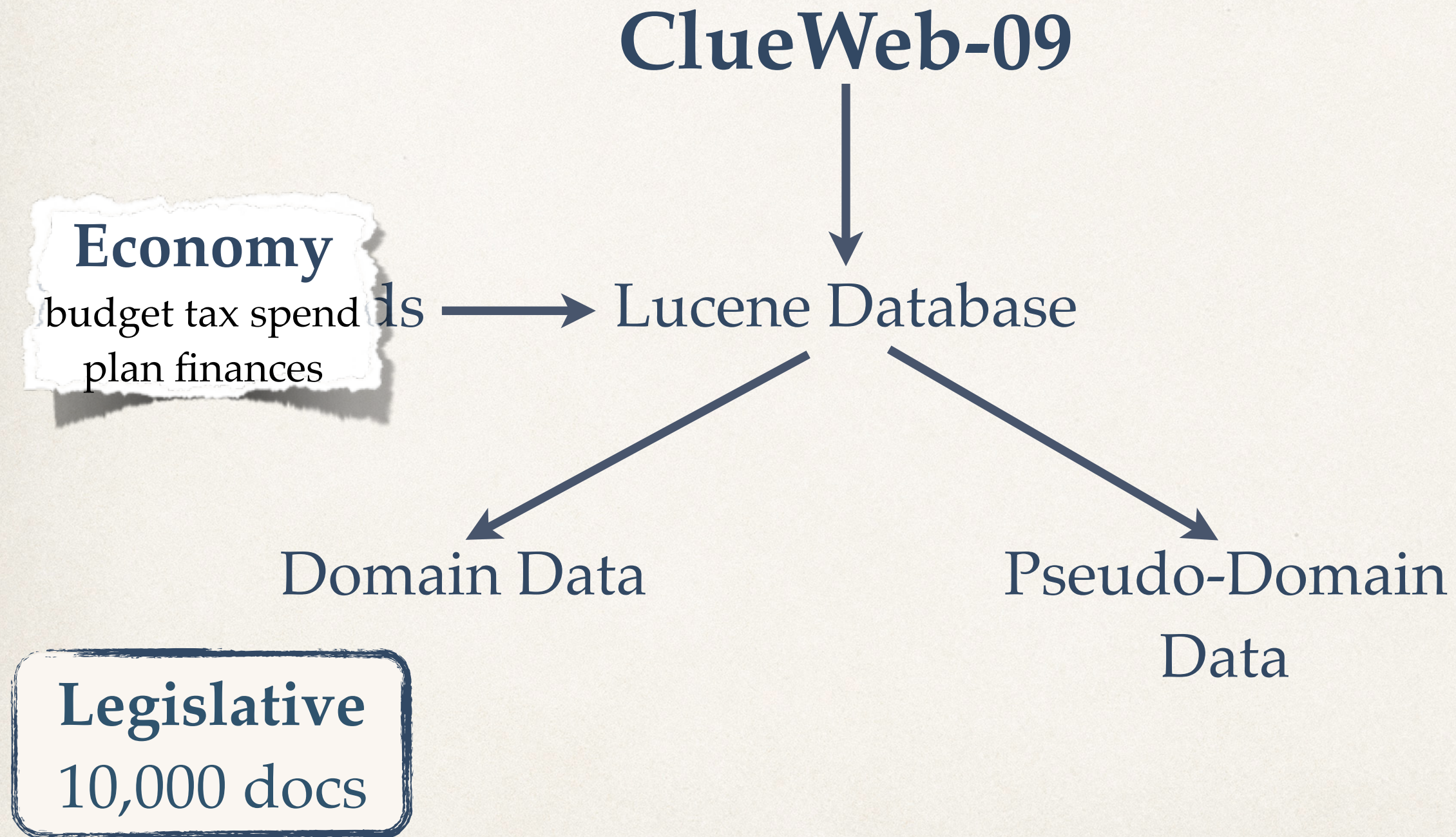
Domain Clustering



Legislative
10,000 docs

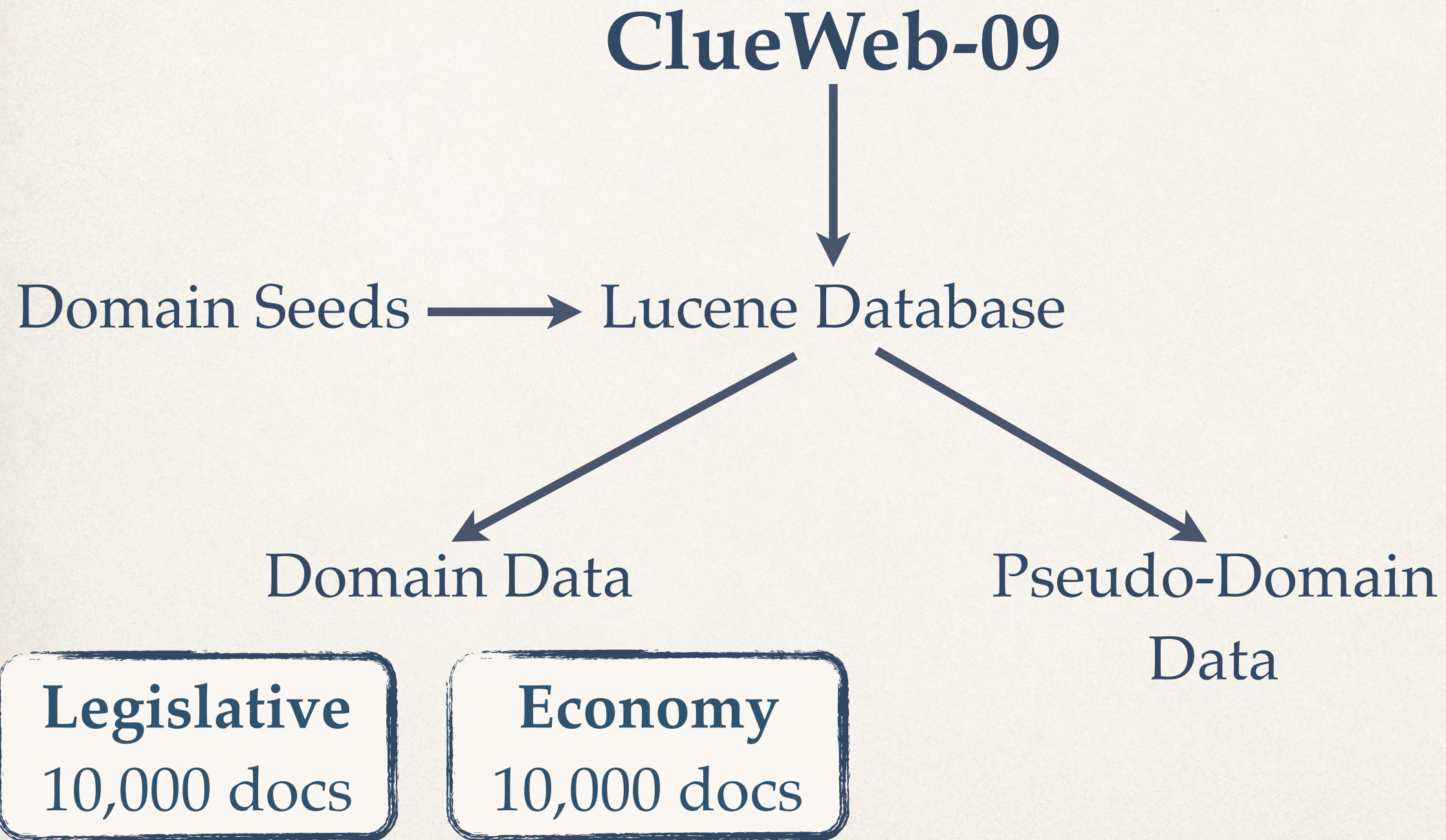
Domain Data

Domain Clustering



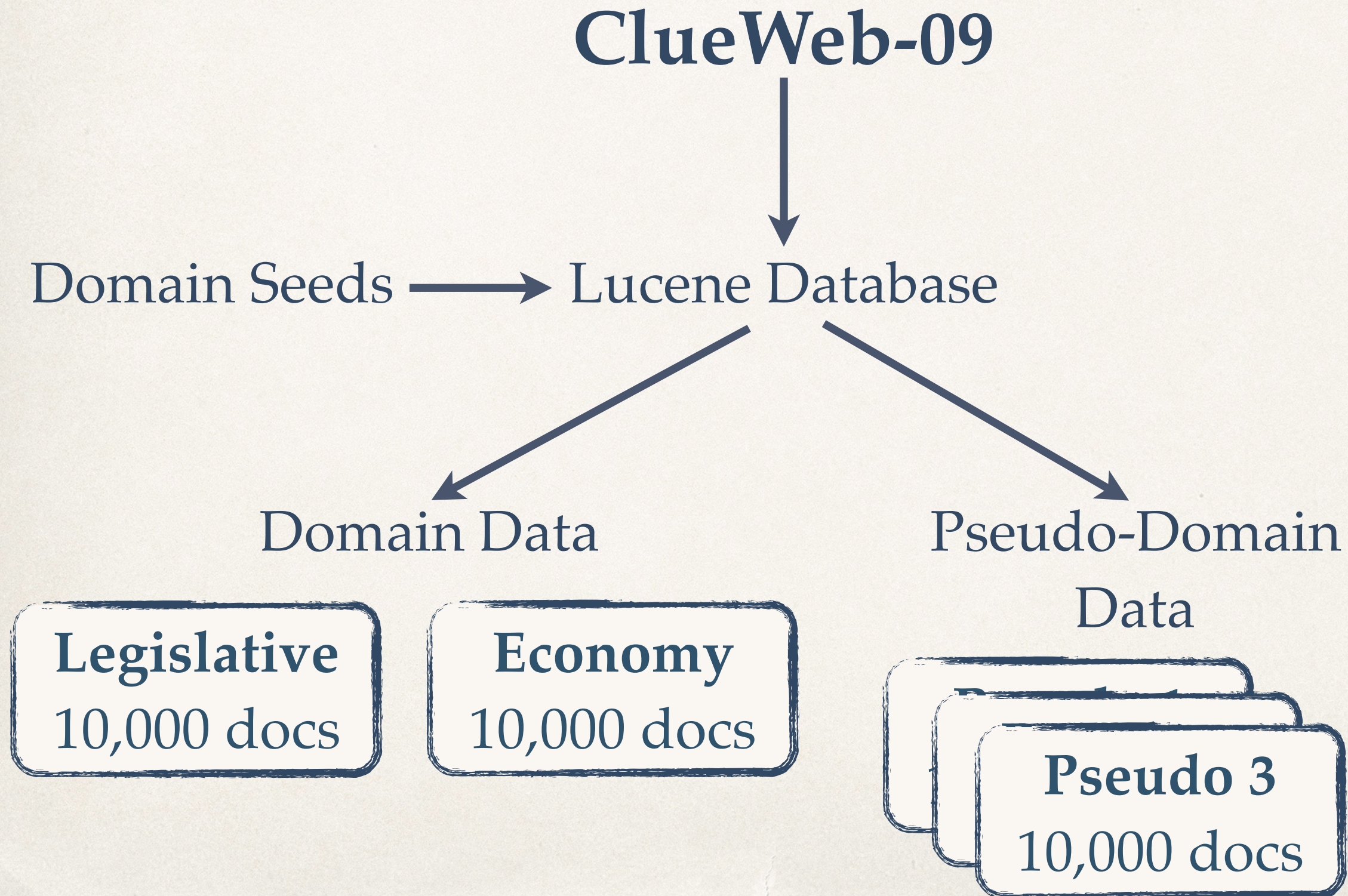
Domain Data

Domain Clustering



Domain Data

Domain Clustering



Overview

Relevance Metric

Domain Data

Evaluation

Basic Classifier

Multi-Feature Classifier

Evaluation

Experimental Setup

“And yet we stand together as we did two centuries ago.”

Experimental Setup

X X X M M X X X X X X

“And yet we stand together as we did two centuries ago.”

Evaluation

Experimental Setup

X X X M M X X X X X X

“And yet we stand together as we did two centuries ago.”

Metrics

- F-Measure
 - Precision
 - Recall
- Accuracy

Evaluation

Experimental Setup

X X X M M X X X X X X

“And yet we stand together as we did two centuries ago.”

Metrics

- F-Measure
 - Precision
 - Recall
- ~~Accuracy~~

Experimental Setup

X X X M M X X X X X X

“And yet we stand together as we did two centuries ago.”

Baseline:

M M M M M M M M M M M

Metrics

- F-Measure
 - Precision
 - Recall
- ~~Accuracy~~

Evaluation

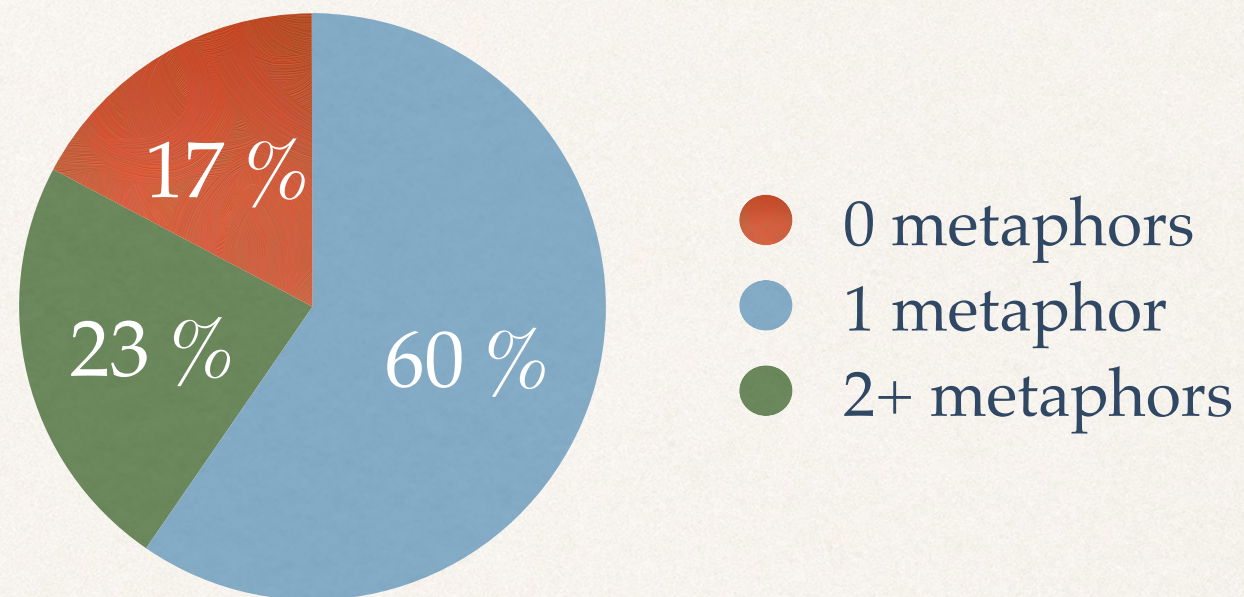
Gold Corpus

MICS Governance Corpus
2510 Sentences

Evaluation

Gold Corpus

MICS Governance Corpus
2510 Sentences

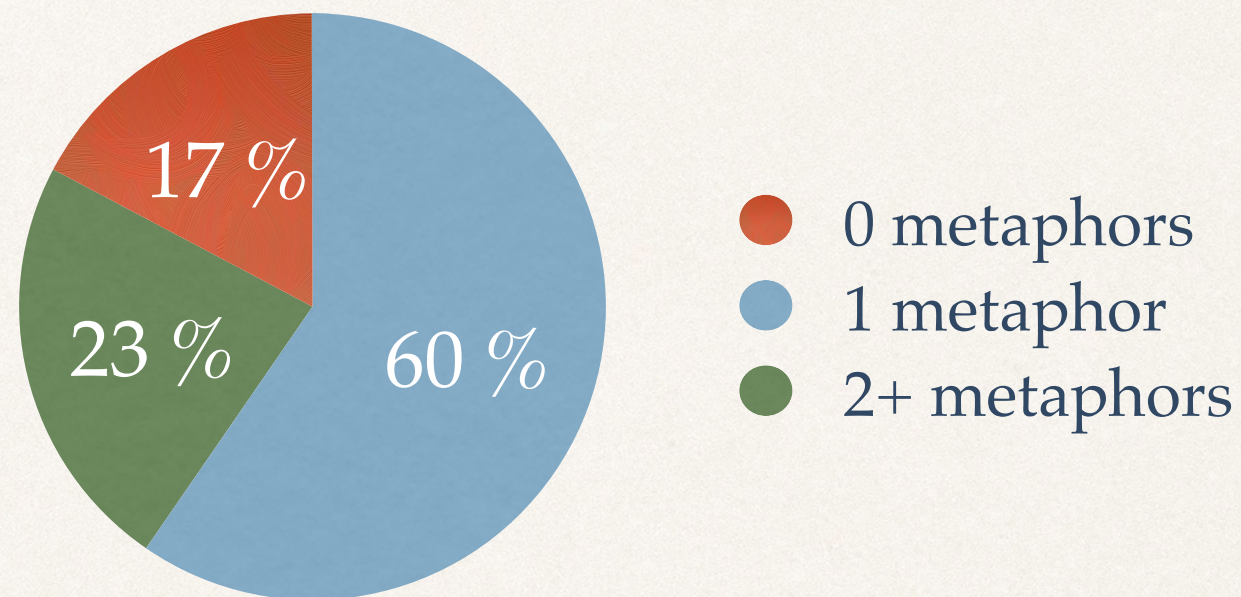


Evaluation

Gold Corpus

MICS Governance Corpus

2510 Sentences



Examples

“And yet we stand together as we did two centuries ago.”

“Many Jewish voters will find themselves at a crossroads.”

Overview

Relevance Metric

Domain Data

Evaluation

Basic Classifier

Multi-Feature Classifier

Basic Classifier

Seeds & Thresholds

Basic Classifier

Seeds & Thresholds

Economy

budget tax spend
plan finances

Basic Classifier

Seeds & Thresholds

Economy

budget tax spend
plan finances

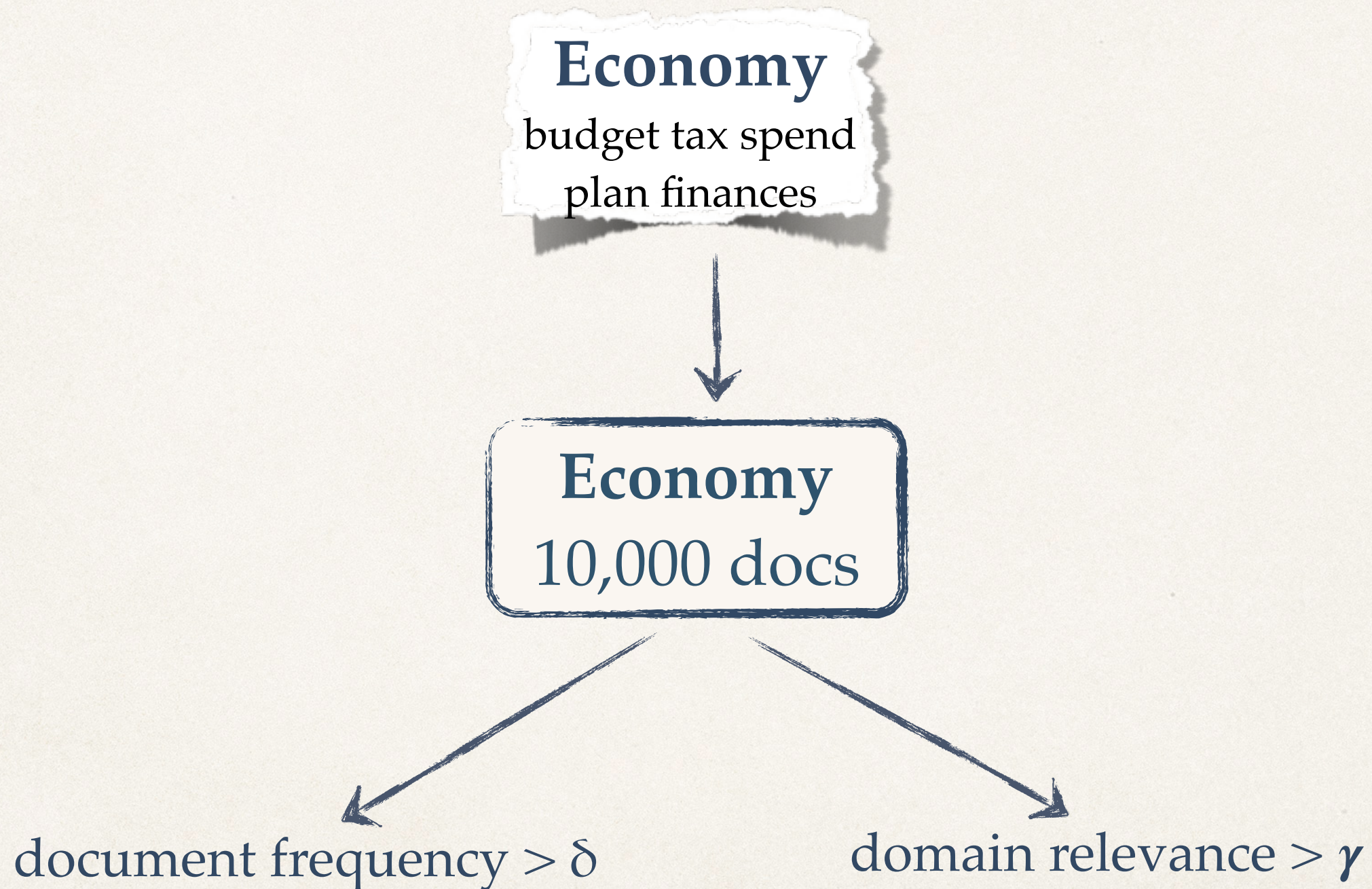


Economy

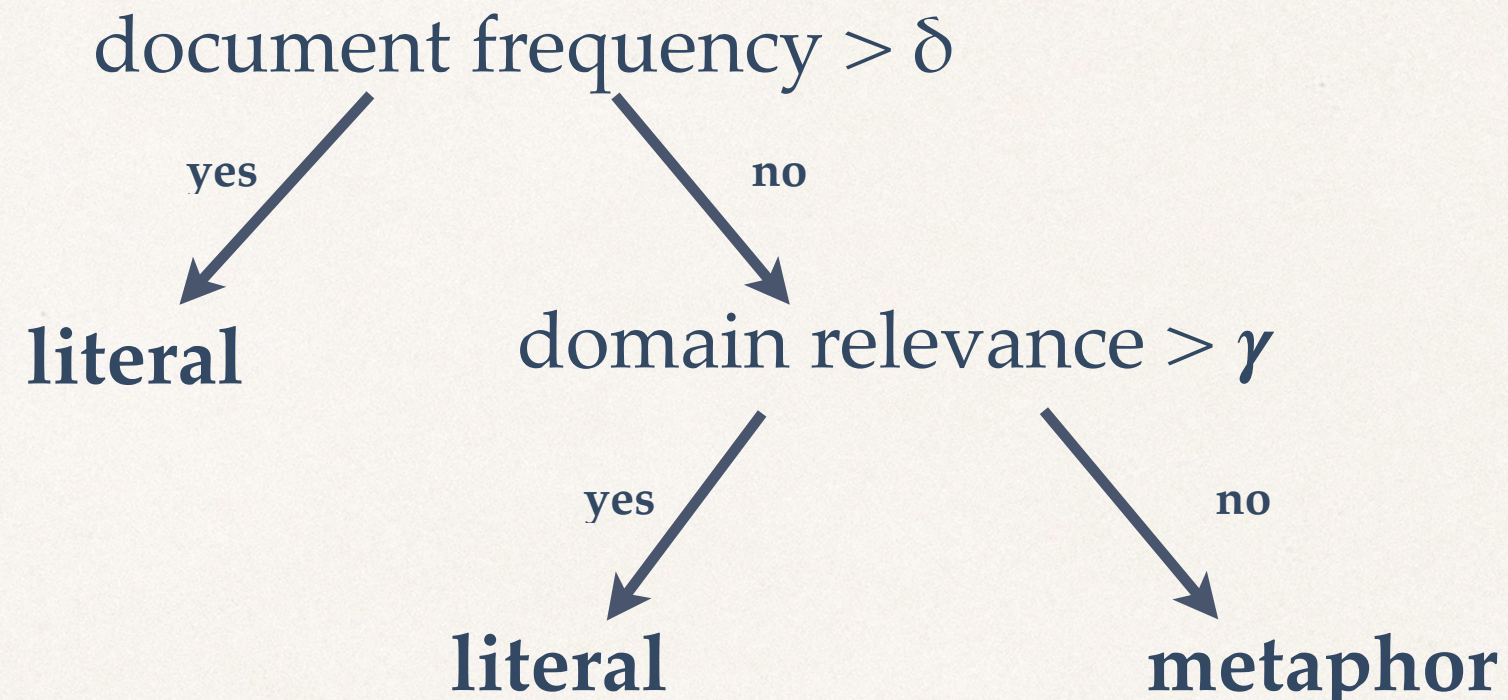
10,000 docs

Basic Classifier

Seeds & Thresholds

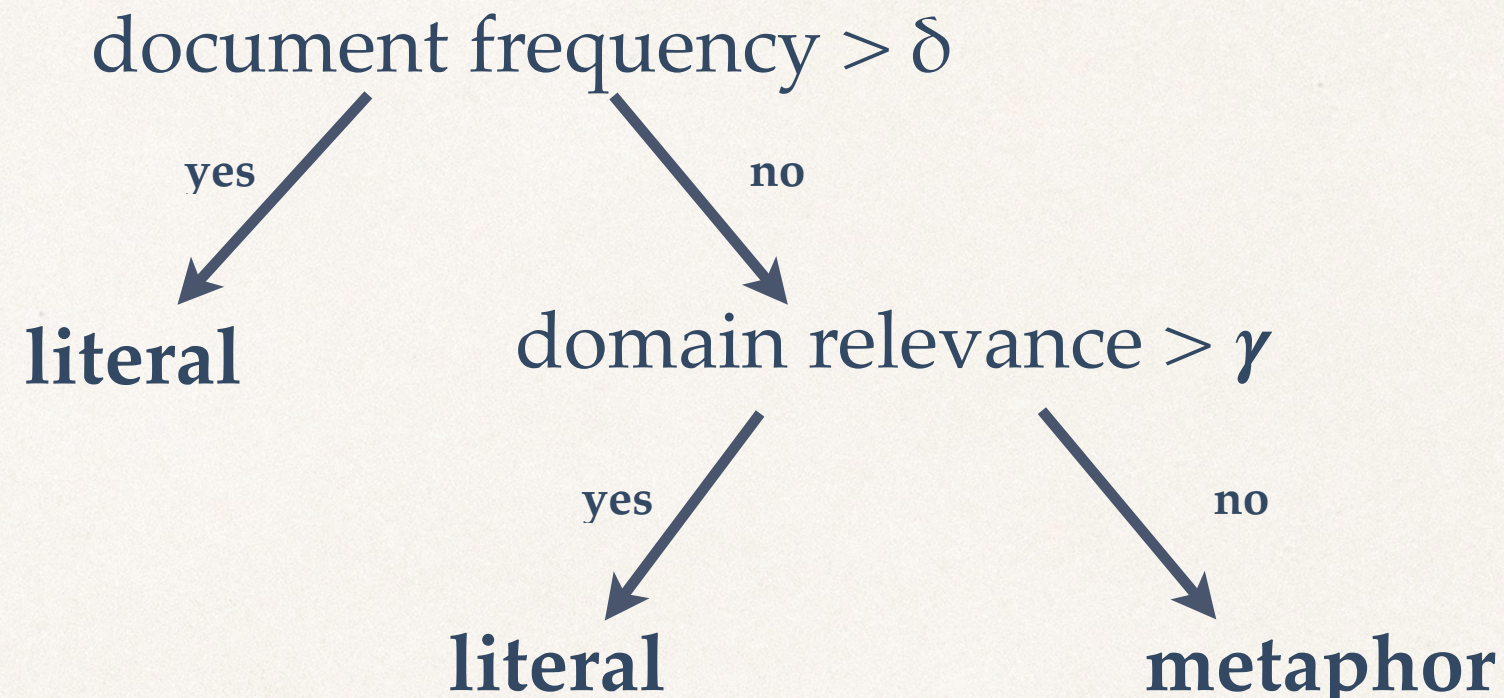


Seeds & Thresholds



Basic Classifier

Seeds & Thresholds



Seed Set 1: Manual

8 Subdomains

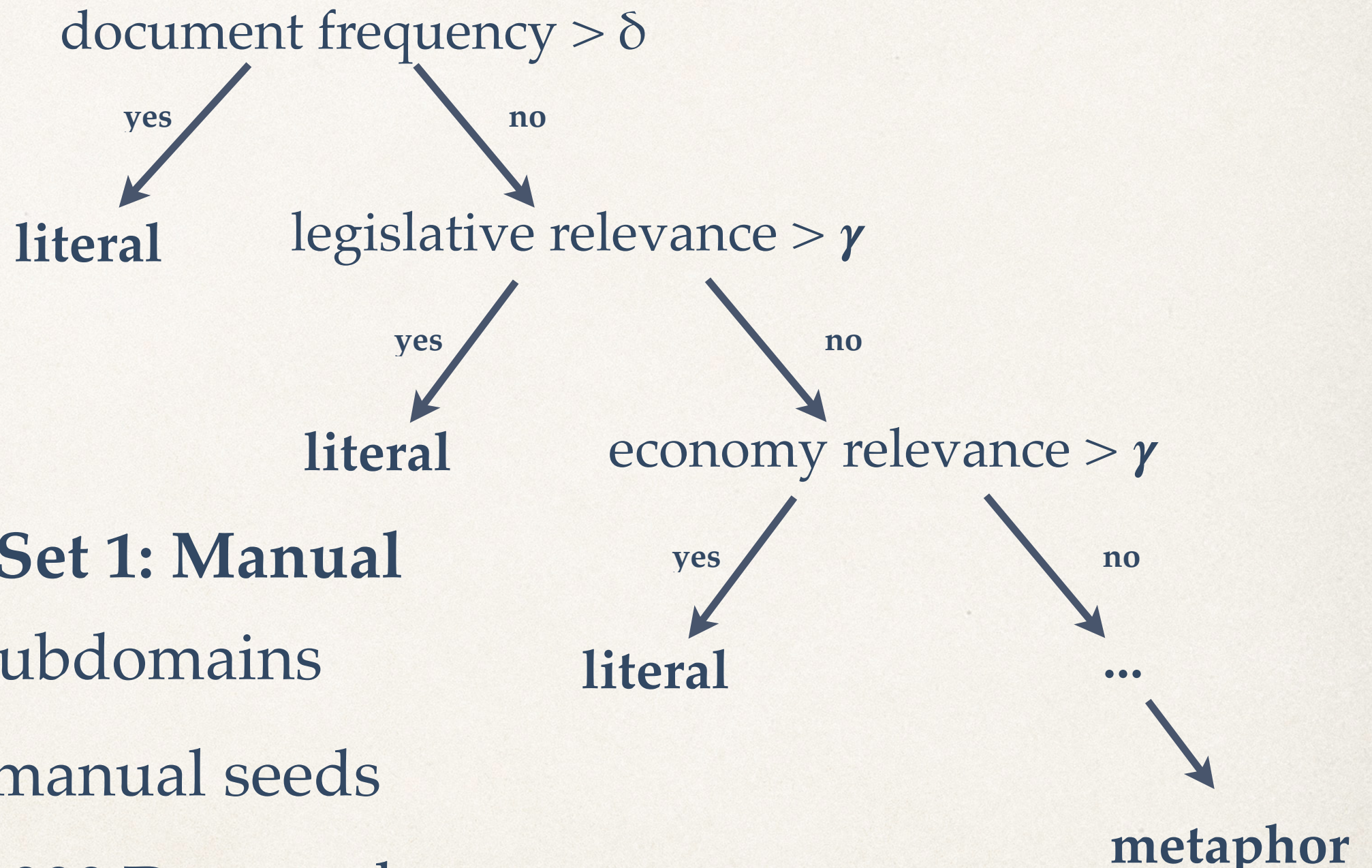
4-14 manual seeds

8 × 10.000 Docs each

$\gamma=0.02$; $\delta=0.1$

Basic Classifier

Seeds & Thresholds



Seed Set 1: Manual

8 Subdomains

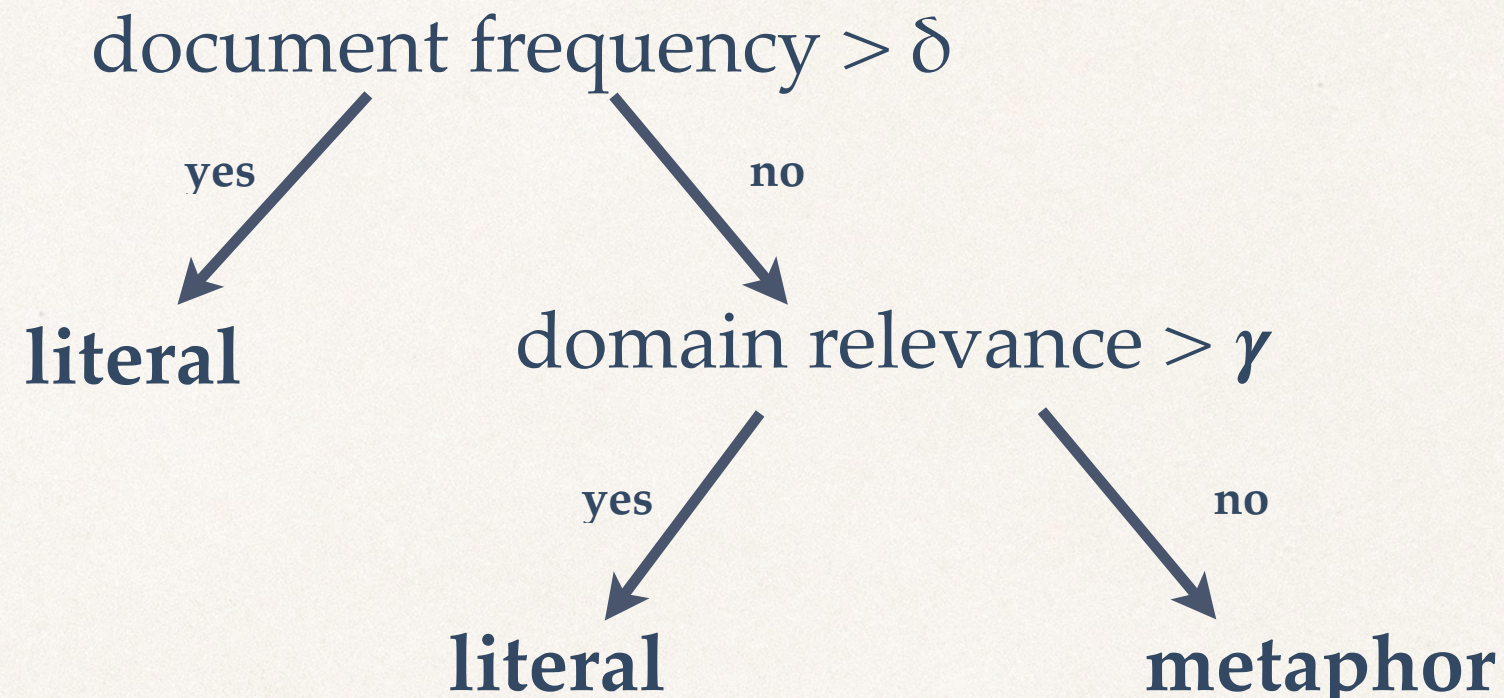
4-14 manual seeds

8 × 10.000 Docs each

$\gamma=0.02$; $\delta=0.1$

Basic Classifier

Seeds & Thresholds



Seed Set 1: Manual

8 Subdomains

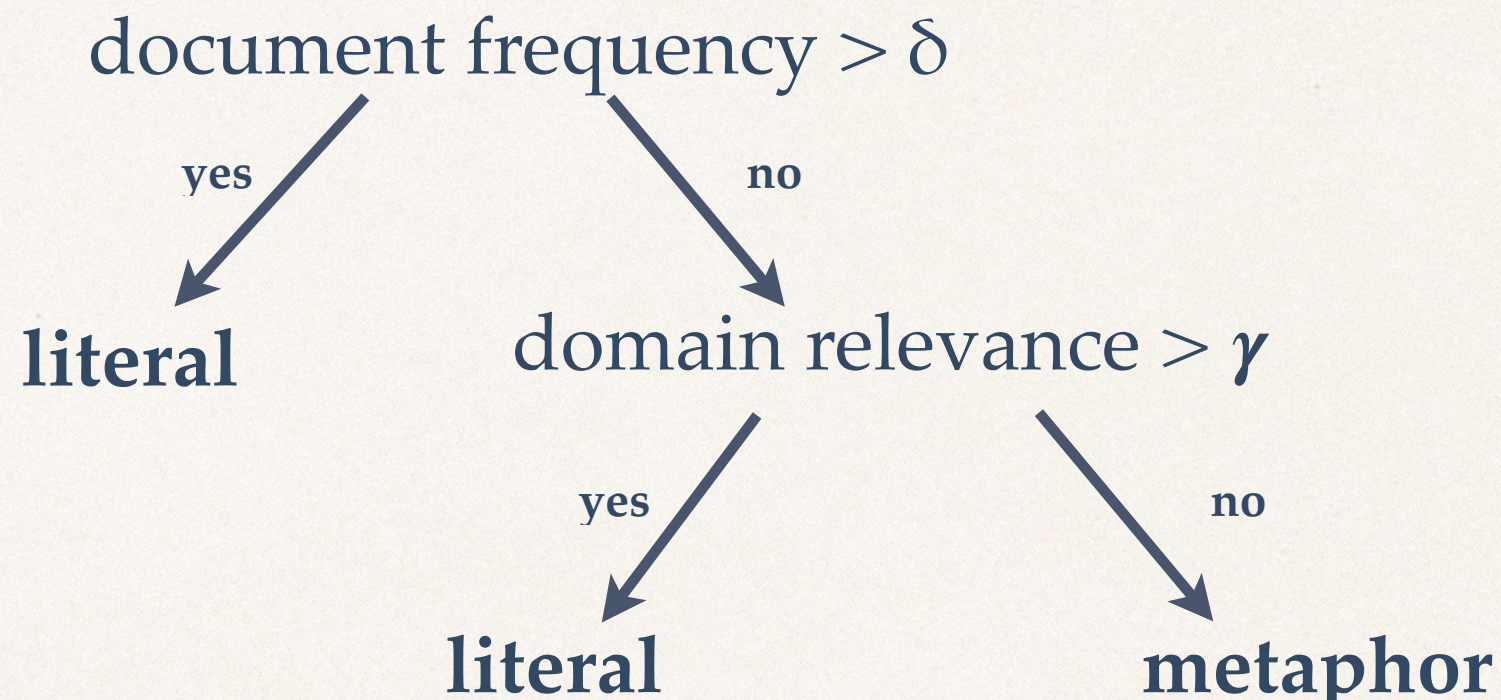
4-14 manual seeds

8 × 10.000 Docs each

$\gamma=0.02$; $\delta=0.1$

Basic Classifier

Seeds & Thresholds



Seed Set 1: Manual

8 Subdomains

4-14 manual seeds

8 × 10.000 Docs each

$\gamma=0.02$; $\delta=0.1$

Seed Set 2: Gold

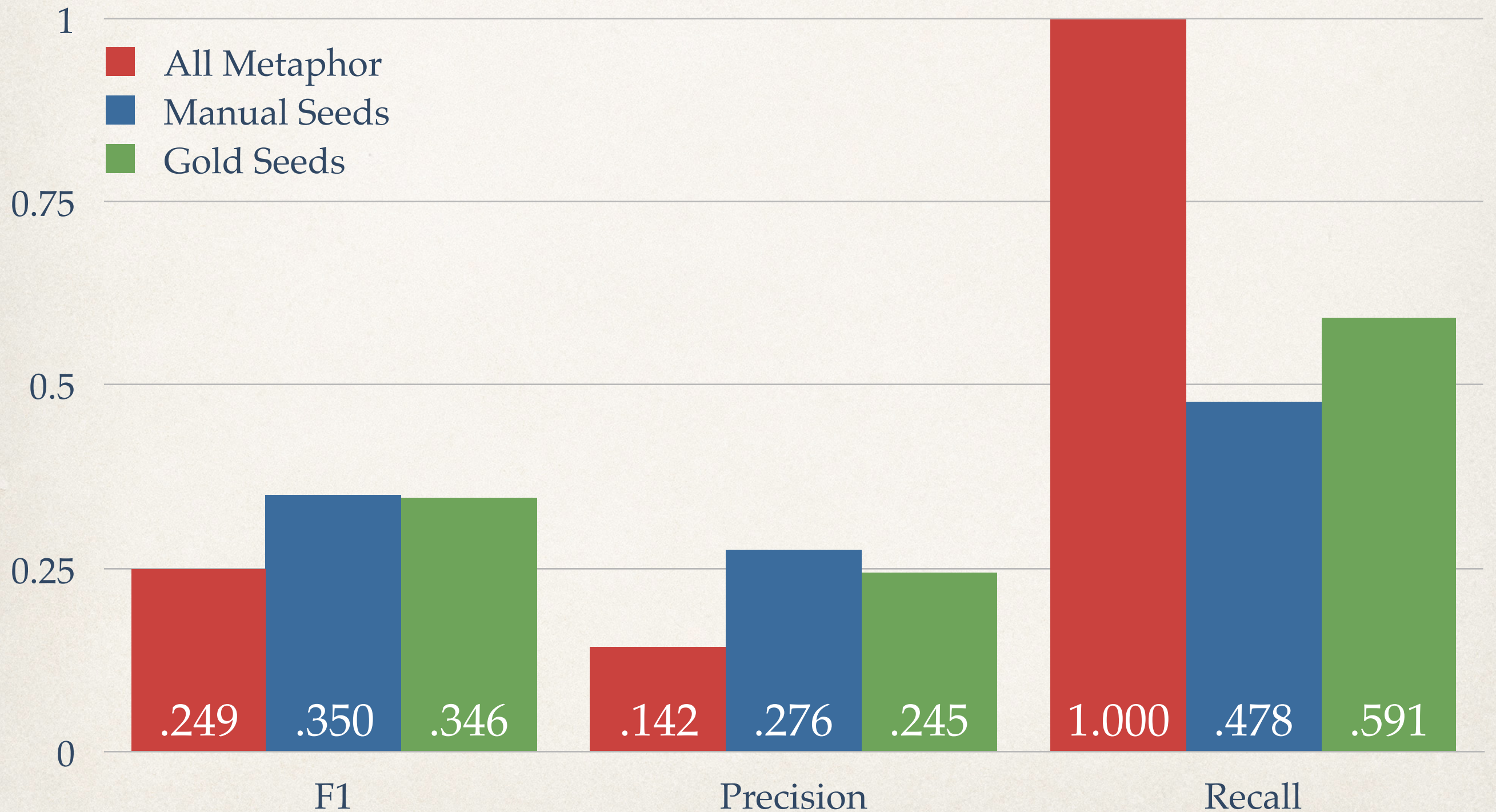
1 Domain

50 best gold metaphors

80.000 Docs

$\gamma=0.01$; $\delta=0.1$

Basic Classifier Evaluation



Overview

Almost Done

Relevance Metric

Domain Data

Evaluation

Basic Classifier

Multi-Feature Classifier

Multi-Feature Classifier

Conditional Random Fields

Setup

Bigram model

10-fold cross validation

Multi-Feature Classifier

Conditional Random Fields

Setup

Bigram model

10-fold cross validation

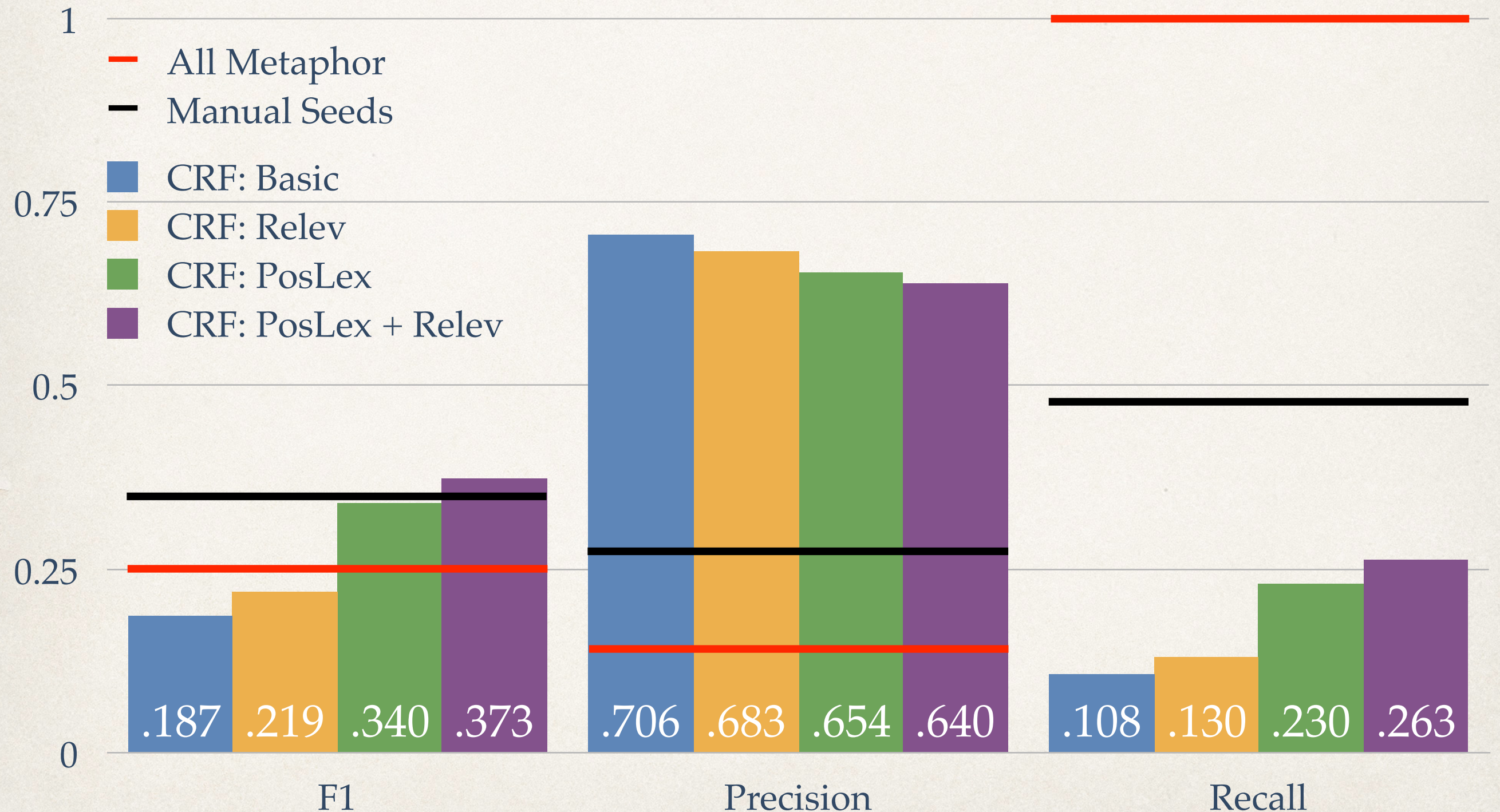
Features

Relevance Weights ($\gamma=0.02$; $\delta=0.79$)

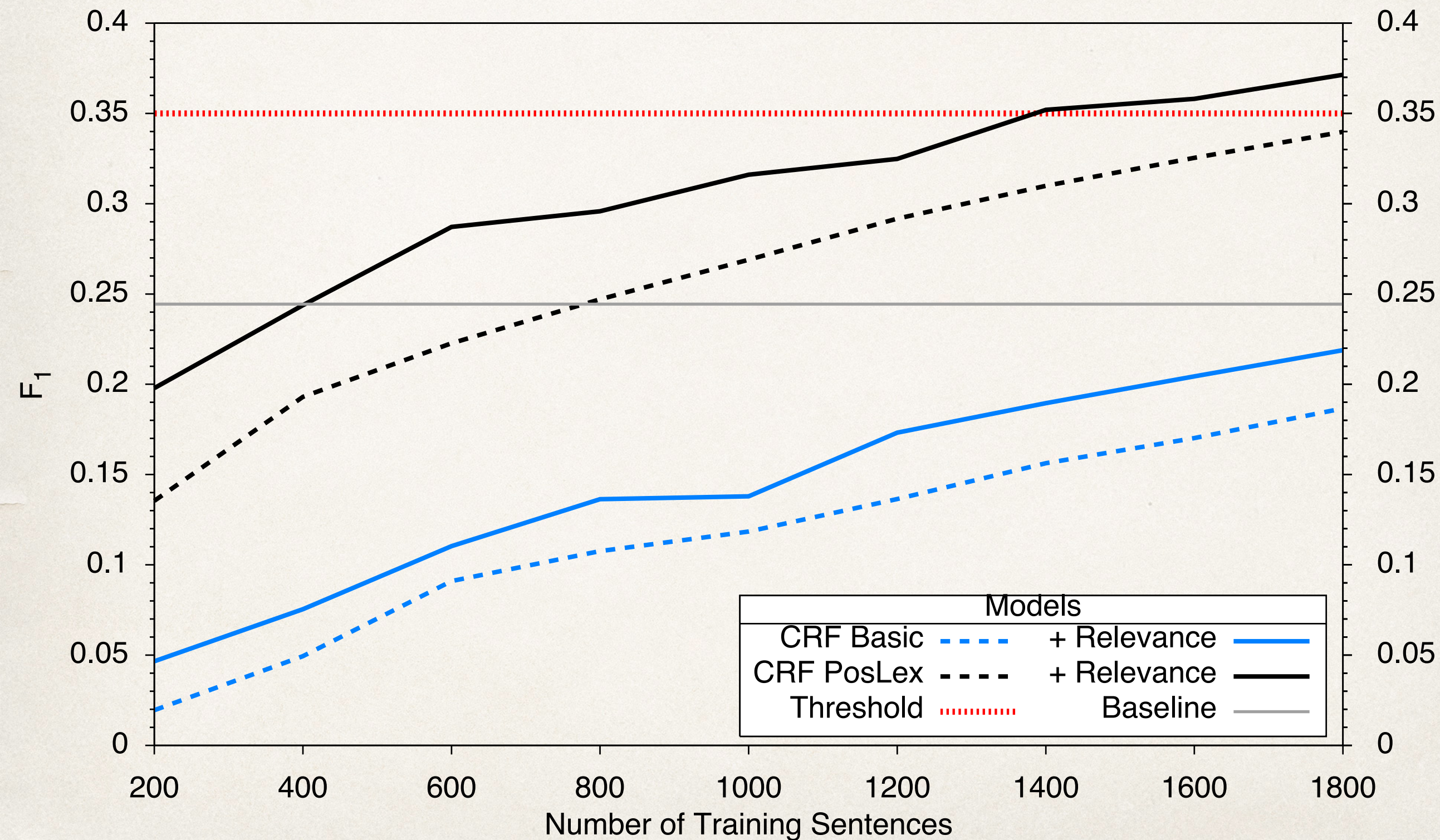
Part of Speech

Lexicographer Sense

Multi-Feature Classifier Evaluation

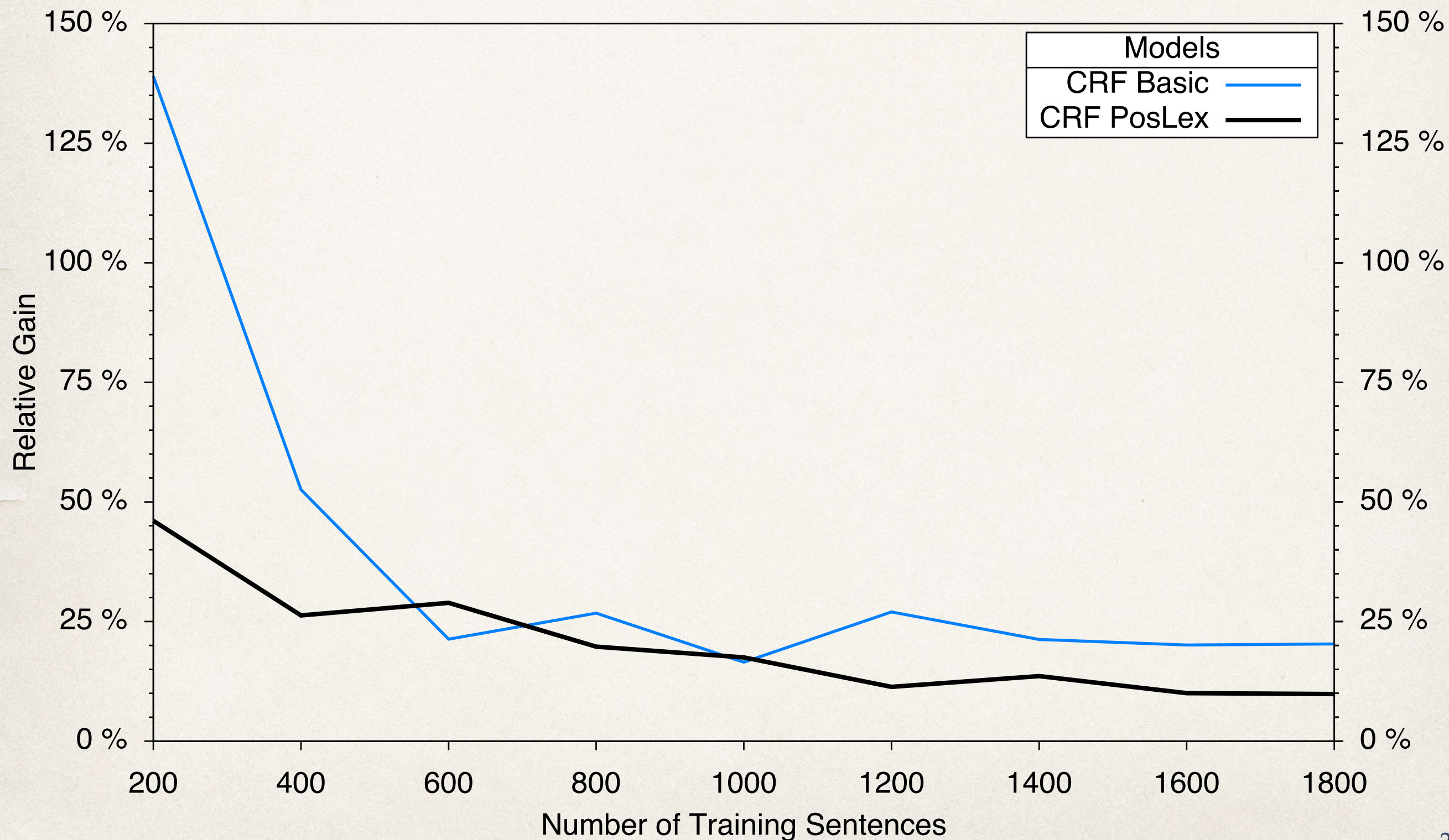


Multi-Feature Classifier Training Size



Multi-Feature Classifier

Relative Gain



Conclusion

Pro

Term Relevance is a cheap Metaphor Heuristic
Useful in Low Resource Scenarios

Con

Performance still too low

Future Work

Term Relevance: Semantic Vector Space

Domain Corpora: Topic Modelling

Application: Other non-literal devices