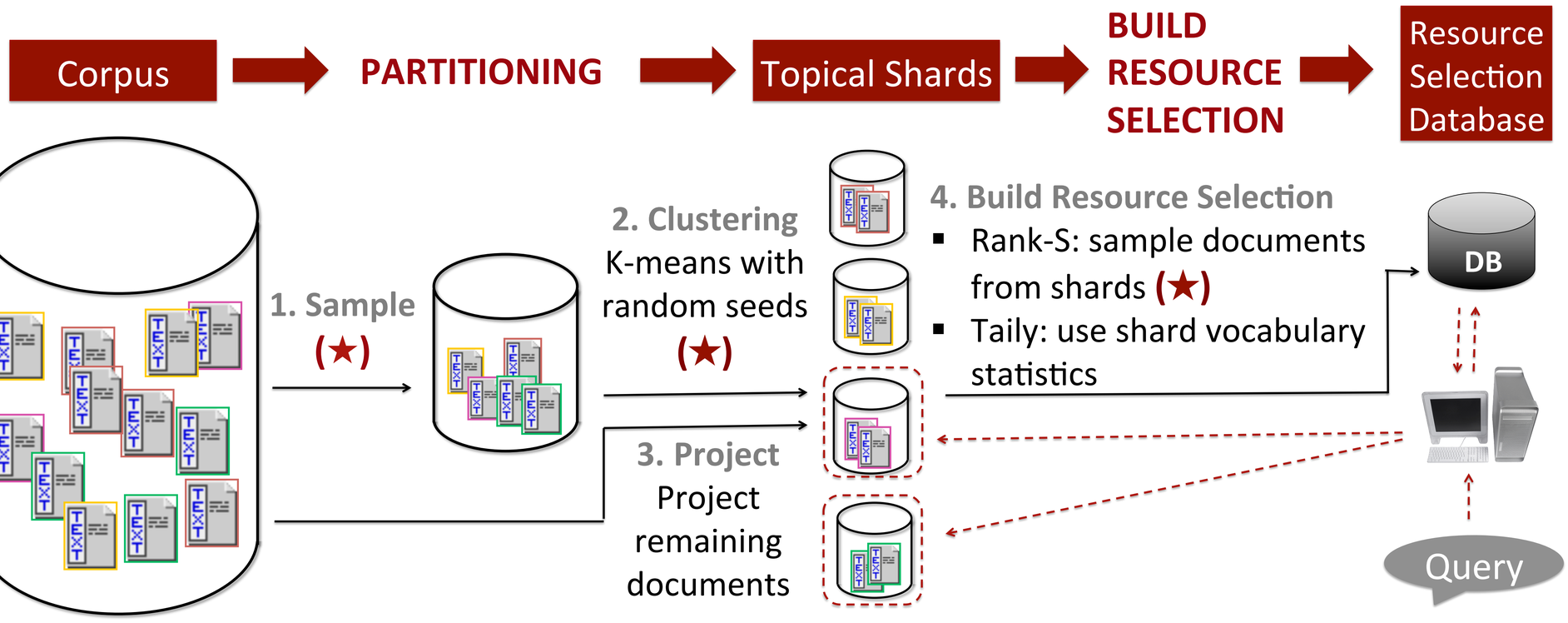
How Random Decisions Affect Selective Search

Zhuyun Dai, Yubin Kim, Jamie Callan
Language Technologies Institute
Carnegie Mellon University

❖ **Selective Search: a distributed search architecture using topical shards to reduce computational costs.**



This work investigated how RANDOM DECISIONS (★) affect variance in search results.

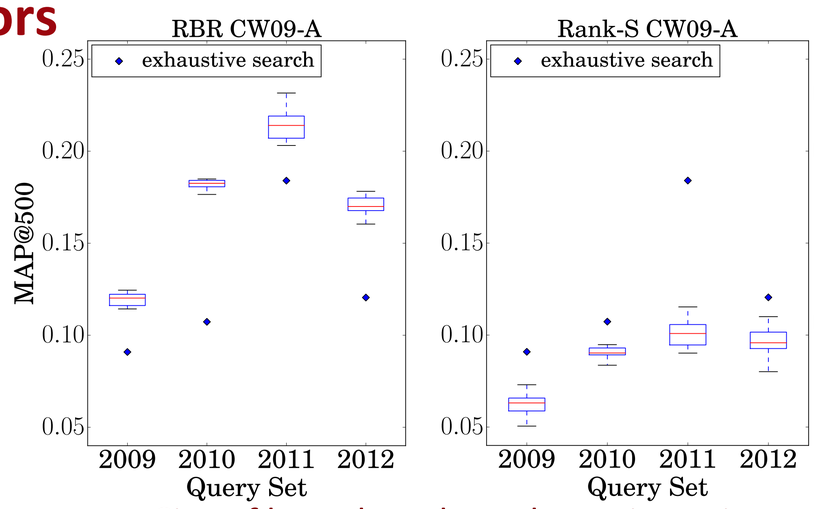
Variability of Query Sets

Experiment Setup

- ClueWeb 09 (Category A, B)
- 4 query sets: TREC 2009-2012
- Resource Selection:
  - Rank-S, Taily
  - **RBR**: ranking-based resource selection. An oracle algorithm.
- 10 Trials

Two Sources of Errors

- **1. Partitioning**
- The only error of RBR comes from partitioning.
- RBR box-plots show that some partitioning results were better than the others.



Size of box-plots show the variance in search results (MAP@500).

- **2. Resource Selection**
- Box-plots of Rank-S were wider than RBR, and had more outliers.
- Resource selection algorithms introduced additional errors due to incomplete models.

Sample-Based vs. Vocabulary-Based

- Rank-S: 3 random decisions
- 2 in partitioning: sampling + clustering
- 1 in resource selection: sampling

- Taily: 2 random decisions
- 2 in partitioning: same as Rank-S
- NO randomness in resource selection

Which is more stable, Rank-S or Taily?

Query Set	Mean of MAP@500		Variance Coefficient	
	Rank-S	Taily	Rank-S	Taily
2009	0.062	0.065	10.65%	10.98%
2010	0.090	0.087	3.66%	6.89%
2011	0.101	0.084	7.62%	15.67%
2012	0.097	0.085	8.23%	5.36%

Mean and Variance Coefficient of MAP@500 of Rank-S CW09-A systems and Taily CW09-A systems.

Variance Coefficient: $\frac{Standard\ Deviation}{Mean}$

- Rank-S has 1 more random component than Taily.
- But Taily only had lower variance than Rank-S on 1 of the 4 query sets.

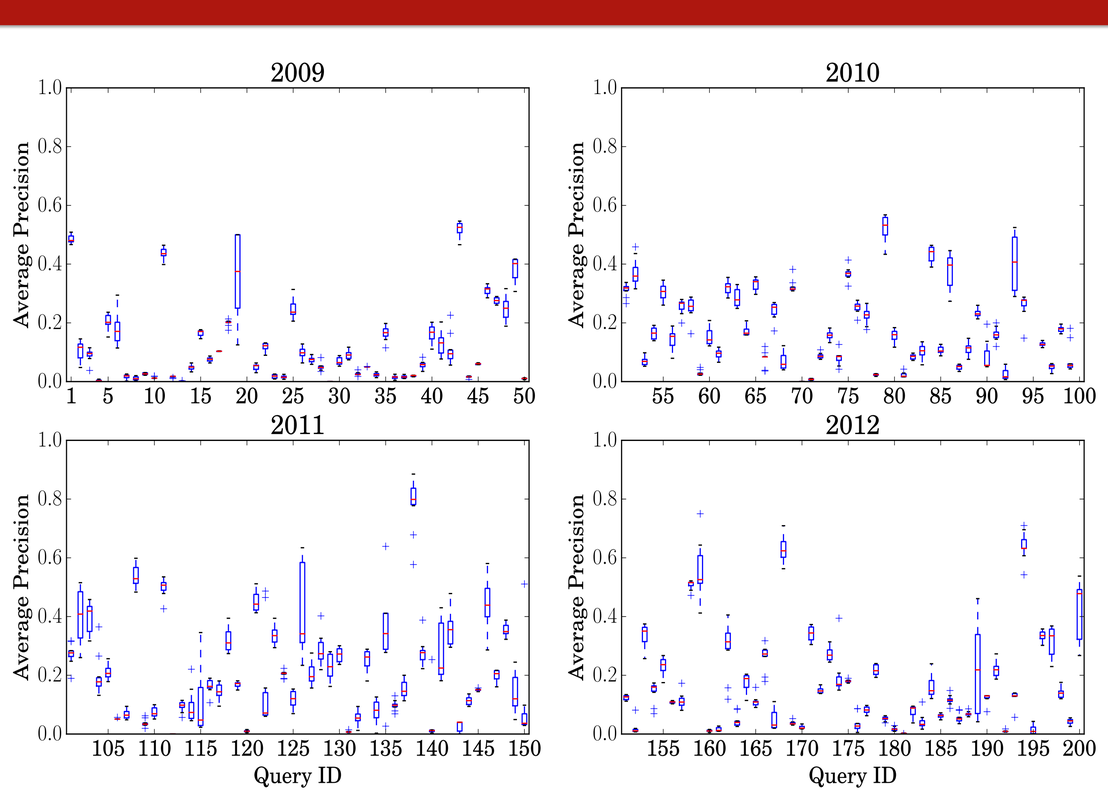
Variability of Queries

Most queries are stable.

- # of high variance queries:
  Rank-S 14 out of 200
  Taily 17 out of 200

Variance mainly comes from partitioning.

- High-variance queries in Rank-S and Taily also had high variance in RBR.
- RBR does not have resource-selection errors. Therefore, the major source of variance comes from partitioning.



Query Average Precision of RBR instances. Each data point is a box and whisker plot with the edges of the boxes representing the upper and lower quantiles, the mid line the median, and the whiskers the 1.5 interquartile range. + are outliers

What is 'poor' partitioning?

- ☒ Is 'poor' partitioning due to relevant documents being grouped into more shards?
- NO. All trials grouped over 60% of relevant documents into 3 or fewer shards.

- ☑ A 'poor' partitioning groups relevant documents with topically dissimilar documents.
- Shards are unrepresentative of relevant documents, making resource selection harder.