

Five Lessons In Designing Cognitive Analytic Systems

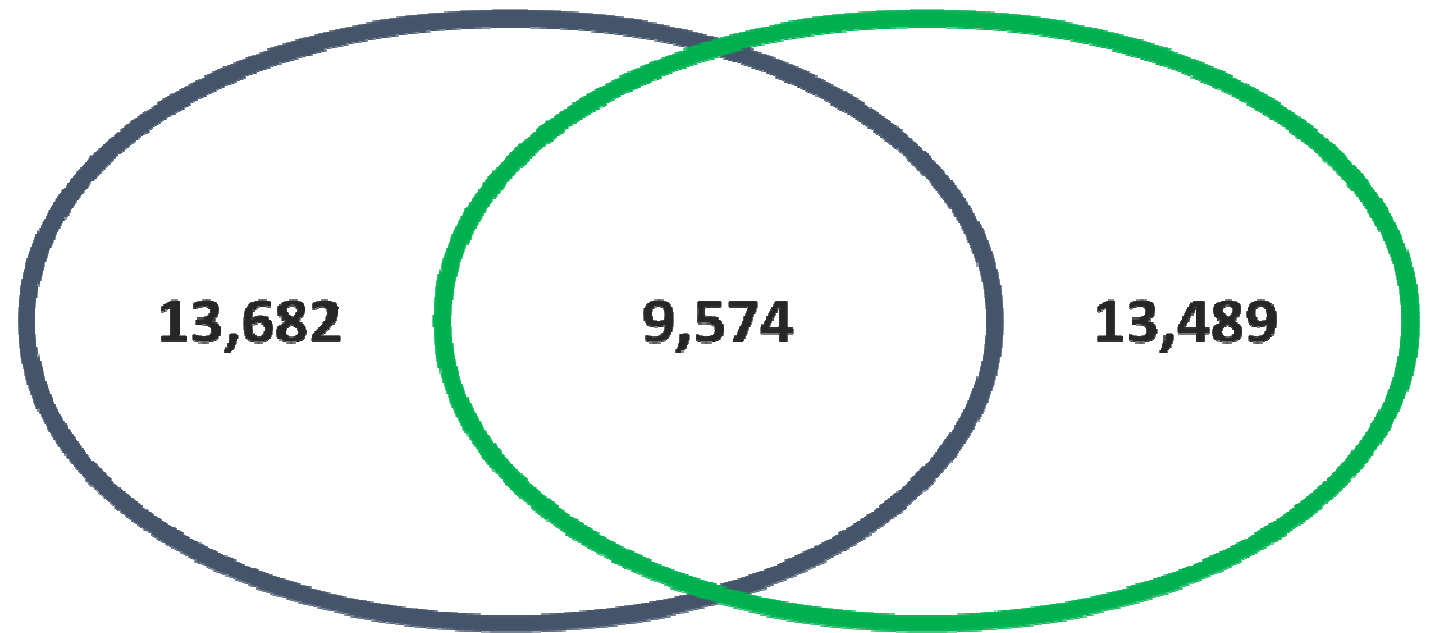
www.coseer.com

1. Be Comfortable With Probabilistic Computing

- Unlike other fields of data science Cognitive Systems give the “most likely” answer.
- Two Platinum standards for Natural Language Processing used for Noun Phrase Extraction on the same corpus had only 70% overlap.
- Coseer uses entire corpuses for deducing summaries or inter-relationships to reduce inaccuracies.

Relative Accuracy

(Number of Noun Phrases found)



Stanford NLP

Ling-Pipe

Recall vs. other

71%

70%

Precision vs. other

70%

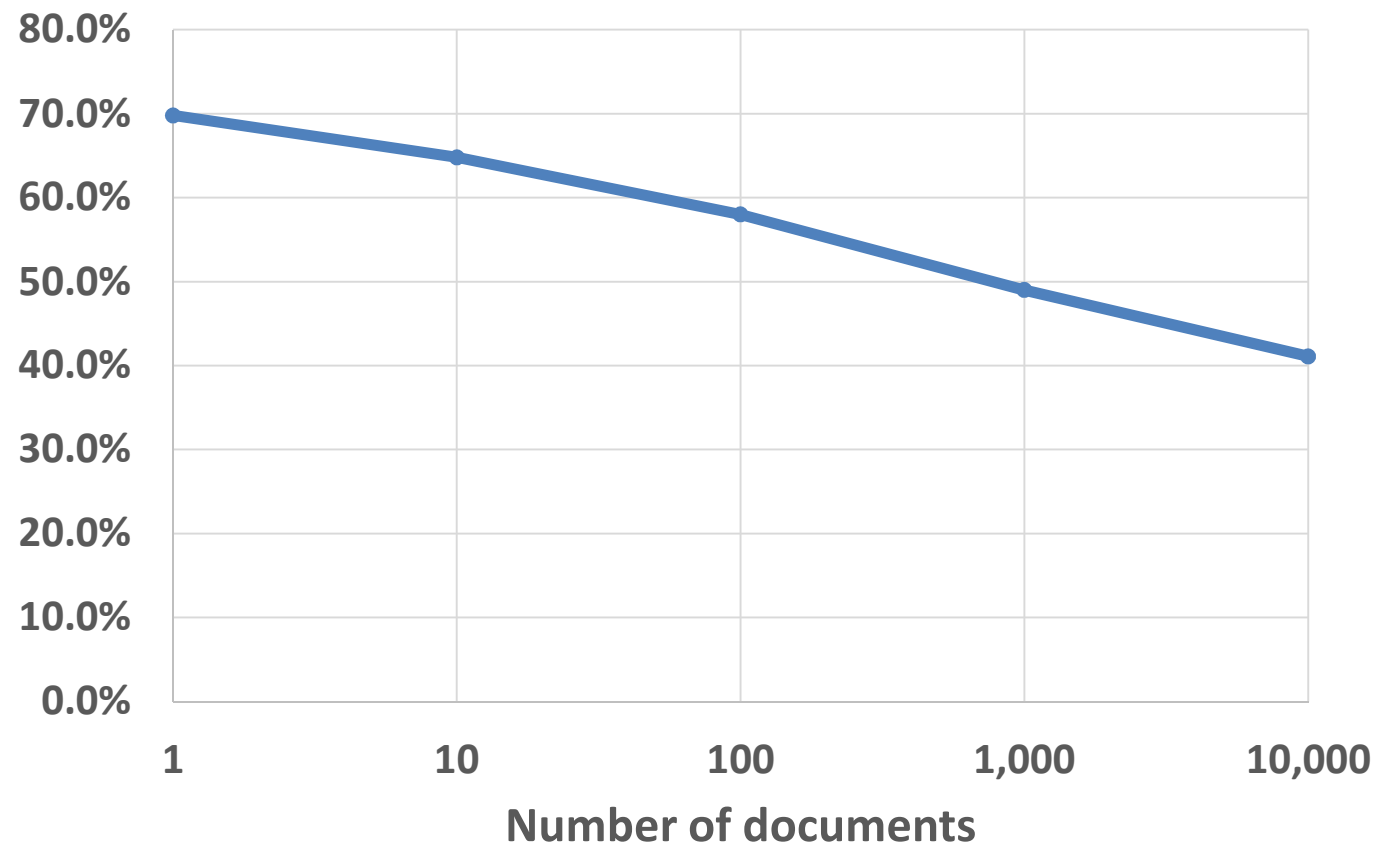
71%

2. Do As Much As Possible In A Single Node

- Completely unpredictable, unstructured, string data
 - Output from each node is similar in size to the input
 - Reduce step requires a lot of computing power.
- Watson is a hardware based solution; Google, Apple, Facebook use Flash DAS.
- Most processes in Coseer are $O(\log n)$ or bounded → there are very few Reduce steps.

Output to Input Ratio

(%, Size of data structures for Noun Phrase Extraction)

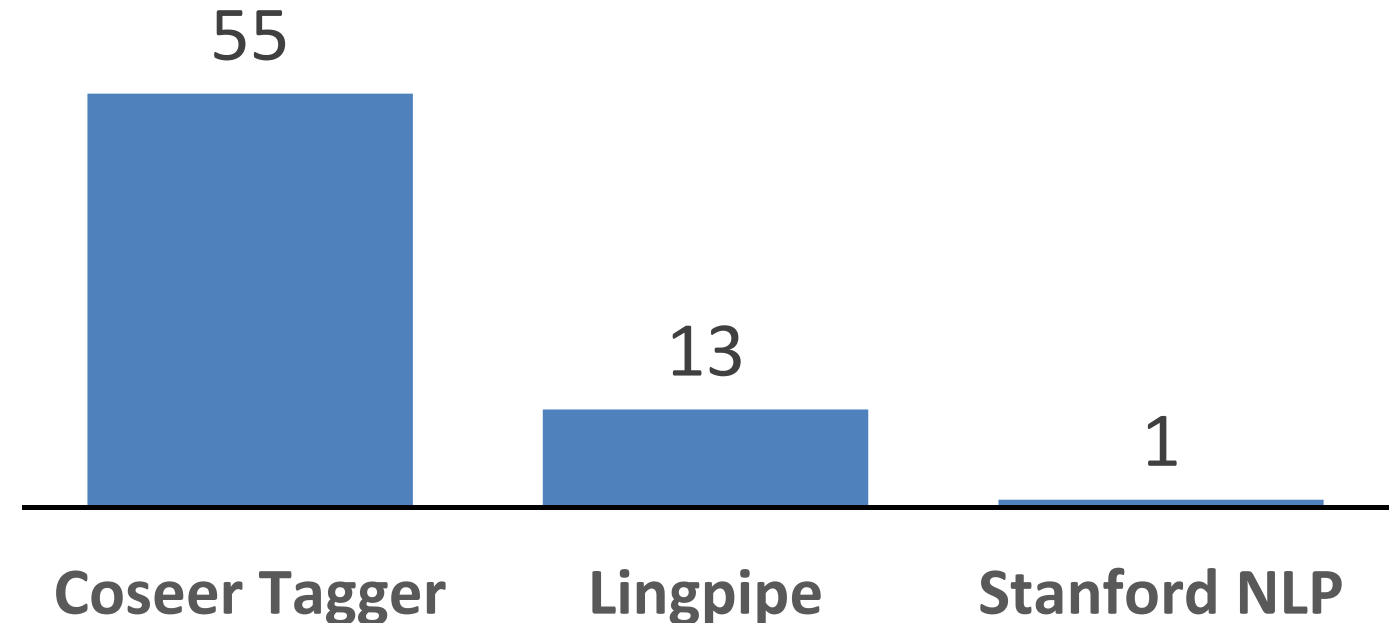


3. Focus on Engineering, Not Science

- Every percent of marginal accuracy may cost exponentially more in terms of computing resources and latency.
- Every percentage of node complexity will exponentially increase complexity of system.
- Coseer traded off ~5% accuracy for 4X throughput.

Chunking Throughput

(Indexed, Stanford NLP = 1)



Vs. Stanford NLP

Vs. Ling-Pipe

Exact Matches

69.8%

64.3%

Matches, where
Coseer finds a
superset phrase

80.2%

89.8%

4. Build Interactive, Context Aware Systems

- Cognitive Analytics Systems are inherently probabilistic
→ Need to involve users in making final decisions.
- Training on specific domains helps with accuracy, but is not sufficient.
- Coseer Express is trained for finance and provides multiple bullets, erring towards higher Recall over higher Precision.



Best Practices

- **Inform, don't decide**
- **Provide multiple options**
- **Give confidence intervals**
- **Encourage feedback**

5. Think Mobile First

- Cliché: Mobile is the future (present?).
- Need for summarization and insights is more acute on-the-go than on a desktop.
- Mobile is a more interactive form-factor.

Sub-topics/ stories
being followed
(automatically curated).

Few bullet points
summarized from 2,286
documents.

Sorted by date, for all of
last week.

Summaries.

Coseer Express
screenshot

