

---

# Cognitive Analytics is Bigger than Big Data Analytics

“Big Data” has been a corporate buzzword for decades. Promising huge efficiency boosts for enterprise search, dazzling customer insights leading to groundbreaking new products and industry-leading customer service, it sounds like the key to running a successful modern business. Cognitive Analytics is a more recent phenomenon, with very similar promises.

Big Data Analytics is primarily concerned with the symbolic processing of vast data for cumbersome problems: enterprise search, targeted advertisements, recommendation systems, and the learning of straightforward correlations. With the advent of scalable, distributed computing, data scientists are now also keen to intelligently manage information that is captured in text, speech and other unstructured forms. An important distinction here is the intelligent management of information. Data management tools often [need experts to constantly oversee](#) their deployment, even supporting text analytics directly. They have unloaded part of the problem to automated systems, but we still need human experts managing these systems.

- What if we were able to train these automated systems to think and learn in the same way as their human managers?
- What if we could structure our data analytics such that anyone could benefit from interaction without intensive training?
- What if we could leverage AI for human language?

Further reading: [What is Cognitive Computing?](#)

Cognitive Computing-enabled Analytics, or Cognitive Analytics, as we call this field, holds much greater potential than Big Data Analytics because it helps unlock the value of Big Data by making the whole system more self-sufficient, and information contained more accessible. It also holds more challenges.

In this post we discuss what it is that makes Cognitive Analytics bigger – and more challenging – than Big Data Analytics.

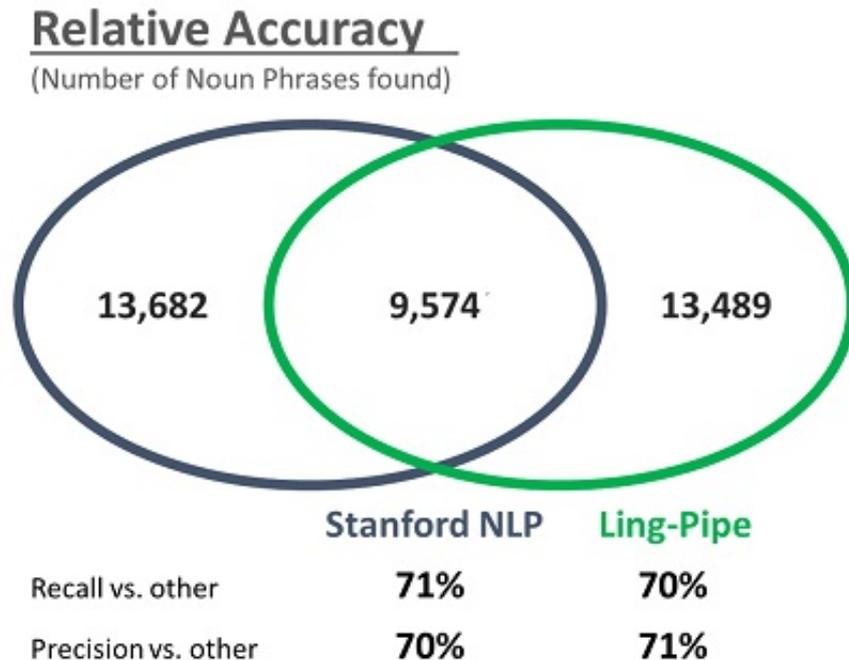
## Probabilistic Computing vs. Deterministic Computing

The answers to analyses conducted using Big Data are usually deterministic. Given a set of assumptions and rules, a machine will give a reliable output. The trick is to get the right set of assumptions and rules, and to program the machine or cluster in a resource-efficient way.

Builders of cognitive systems are not so lucky. Cognitive Analytics deals with problems that are inherently uncertain; problems whose solutions are unavoidably and necessarily probabilistic.

---

Consider a classic probabilistic problem: natural language processing (NLP). [Stanford NLP](#) and [Ling-Pipe](#) are two of the leading software packages in this area: both claim 98% accuracy, but if you ask them to process real-world data, the probabilistic nature of language means they only agree 70% of the time. We know because we tested this in 2015, asking both Stanford and Ling-Pipe to extract the noun phrases from a corpus consisting of one week's content posted by a popular Wall Street Twitter account.



Leading NLP solutions agree with each other only 70% of the time for something as basic as Noun Phrase Extraction.

What our Wall Street experiment demonstrates is we can only be 70% confident that any noun phrases extracted by industry-standard NLP tools are accurate.

Coseer ameliorates this problem by using Noun Phrase Chunking, a Cognitive Analytics technique that identifies key topics in a corpus and uses them to learn associations between data. This creates more data points and true 98% accuracy, so analysts can be confident output data is trustworthy. The answer itself of course is still necessarily probabilistic, but now underwritten by extensive corpus-based processing.

Further Reading: Learn more about how Coseer extracts [meaningful output from unstructured data](#).

---

## Cognitive Analytics is Difficult to Map-Reduce

Unlike Big Data Analytics, most Cognitive Analytics tasks do not render themselves to the map-reduce style of distributed computing. Due to completely unpredictable, unstructured and string-based data the computational complexity of the reduce step tends to be the same order of magnitude as the complexity of the original problem.

For a simple problem of finding the phrase from a corpus with the highest TF-IDF, each node in the cluster will output a data-structure that is almost as bulky as the input corpus – only a very powerful node, therefore, is able to complete the reduce step.

### What does this mean for the the AI Industry?

**Ergo, Cognitive Analytics systems should be designed to execute as much as possible in a single node and use problem-specific distributed architecture where needed. That is partly why IBM Watson is a hardware-based solution, a derivative of the Blue Gene supercomputer. This is also one of the reasons Google, Facebook, Apple, and other companies dealing with large quantities of text, design their data centers using super-fast, flash-based Direct Attached Storage rather than prevalent network storage (SAN or NAS).**

Coseer's Cognitive Analytics technology distills relevant information from disparate sources and presents it in an easy-to-use format that improves productivity, saves valuable time and resources, and reduces response times for enterprises and individual users. [Click here to read more about our design philosophies, and why we use Tactical Cognitive Computing.](#)

Coseer takes an engineer's approach to the problem. For the step of Noun Phrase Extraction, we compare Coseer to the platinum standards in the image below.

<u>Coseer Recall</u>		
	Vs. Stanford NLP	Vs. Ling-Pipe
Exact Matches	69.8%	64.3%
Matches, where Coseer finds a superset phrase	80.2%	89.8%

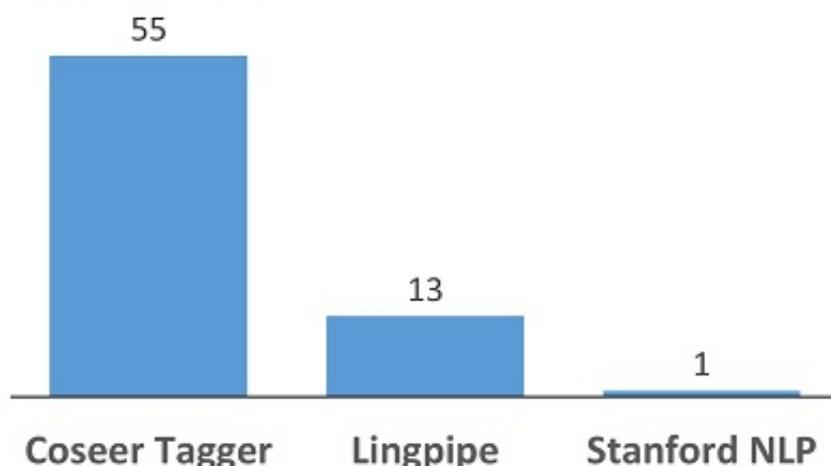
An engineering oriented approach does not compromise on quality of output.

What we give up marginally in accuracy, we make up with our corpus-based approach. As well as being incredibly memory-efficient, our technique is also far superior in terms of time taken, as demonstrated in the next graph.

---

## Chunking Throughput

(Indexed, Stanford NLP = 1)



Given that we need to process billions of sentences, this superiority becomes a defining factor in our solutions. Systems without such innovation have to trade-off accuracy if they are to manage the complexity of cognitive analytics.

### The Need For Context Aware, Interactive Design

Ultimately, analytics are meant to enable better business decisions through effectively leveraged data. To this end, many Big Data Systems are employed to make decisions once trained. Due to the probabilistic nature of tools available, Cognitive Analytics Systems must be designed for higher interaction. This is true not only for learning, but also for decision-making. High levels of interactivity increase a user's confidence in the system, improve the accuracy of results and reduce the complexity of challenges. Such interactivity also provides the context for further improvement. There is enough mystery when talking about AI - large enterprises with vast, valuable data sets need an AI that isn't a black box.

Similarly, the output from Cognitive Analytics should come with attached confidence intervals. It is also possible for such systems to report multiple options and let the user choose which option is best. Such considerations increase the challenge for software designers to include features for easy interactivity.

### What next?

**Like Big Data Analysis Systems, Cognitive Analysis Systems are more effective and accurate when domain-specific information is used in training so that the inference is tuned to the specific structure and needs of the problem under consideration.**

**Cognitive Analytics, as a field, is developing fast, and a lot more exciting things are yet to come. If you're interested in how this all works, if you have a specific data set(s)**

---

you'd like to try out, check out our [introductory video](#) on Coseer's capabilities, set up [a meeting with our tech team](#).

## What is Next-Generation Enterprise Search?

Coseer's search solutions are transforming industries from healthcare to finance. Our point-and-shoot AI trains finds answers and insights with 95%+ accuracy within 4-12 weeks - all of this in 100% security. The reason? We founded Coseer on the principle that computers should take care of the boring stuff so that humans can focus on creativity and judgment. To that end, we've built enterprise search solutions to complete complex workflows just as humans would in a fraction of the time. Fortune 500 leaders are using Coseer to speed up and automate their most complex work.

We follow a tactical approach to enterprise search:

- We deliver 95-98% accurate solutions within 4-12 weeks.
- Our solutions deploy entirely behind your own firewall for 100% security, and every decision point is logged for full transparency.
- You add the finishing touches, but our point-and-shoot AI practically trains itself. No more huge training data sets or time wasted annotating and tagging.

Visit our [website](#) for in-depth case studies, ROI breakdowns per industry, and other insight.