What Is NLP?

Introduction to Natural Language Processing

What Is NLP?

- In computer science, NLP typically refers to the development and use of machine-based methods to process content in *natural language.*
- So what is meant by a “natural language”?
- We say that to distinguish it from an artificial language. What’s the difference between those?

NLP

From Wikipedia, the free encyclopedia

NLP may refer to:

Computer science

- Natural language processing, the field of computer science concerned with human speech as it is spoken.
- Natural language programming, an ontology-assisted way of programming in terms of natural language sentences.

Networking

- Normal link pulses, a signaling mechanism used in Ethernet

Political parties

- National Labour Party
- National Liberal Party
- New Labour Party
- Natural Law Party

Libraries

- National Library of the Philippines

Personal development

- Neuro-linguistic programming

Natural vs. Artificial Languages

- Natural languages are those that evolved or emerged gradually over time, largely unconsciously.
- Artificial languages are those that were designed, crafted, or invented with conscious purpose, largely all at once and not gradually.

Morgenstund hat Gold im Mund.

The early bird catches the worm.

vs.

```
for i := 0 to NumPhrases do
  if Pos(Phrases[i],CombinedStr) <> 0 then
    Score := Score+((1-Score)*DTG_Bonus);
```

Examples

Every person in this class is competent in at least one natural language and at least one artificial language—the same two languages, shared by everyone in this class.

Can you name them?

Examples

Every person in this class is competent in at least one natural language and at least one artificial language—the same two languages, shared by everyone in this class.

Can you name them?

- That's right: English and 

Examples of Natural Languages

- English
- Chinese
- German
- Japanese
- Latin
- Hebrew
- Greek
- Russian
- Sanskrit
- Arabic
- French
- Italian
- Spanish

Examples of Artificial Languages

- LISP
- Prolog
- C/C++
- Java
- Javascript
- Scala
- Python
- Perl
- Pascal
- Ruby

Examples of Artificial Languages

But wait!

Those are all programming languages.

Does “artificial language” just mean “programming language”?

No. There are other types of *very famous* artificial languages. Can you think of some? We’ll go over that in the next segment.

DataScience@SMU

Natural vs. Artificial Languages

Introduction to Natural Language Processing

Natural vs. Artificial Languages

- English
- Chinese
- German
- Japanese
- Latin
- Hebrew
- Greek
- Russian
- Sanskrit
- Arabic

- LISP
- Prolog
- C/C++
- Java
- Javascript
- Scala
- Python
- Perl
- Pascal
- Ruby

Examples of Artificial Languages

Can you think of some more artificial languages that are not computer languages?

Examples of Artificial Languages outside of Computer Programming

Fiction:

- Elvish (J. R. R. Tolkien)
- Klingon (Marc Okrand)

Examples of Artificial Languages outside of Computer Programming

International communication:

- Esperanto (L. L. Zamenhof)
- Interlingua (Alice Vanderbilt Morris)

Examples of Artificial Languages outside of Computer Programming

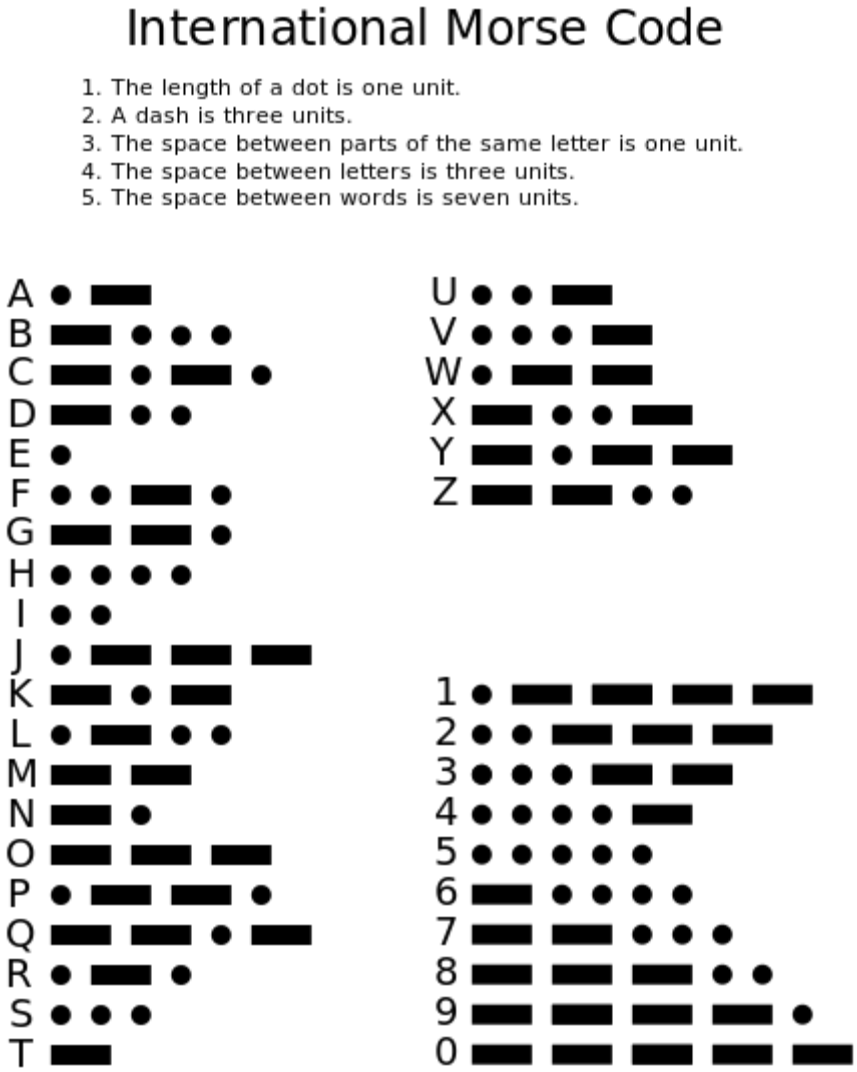
Language games (or for secrecy):

- Pig Latin (based on English)
- Verlan (based on French)

Clarify the Distinction

Q: Is Morse code an artificial language?

A: No, because it is just a code for the alphabet and numerals.



Clarify the Distinction

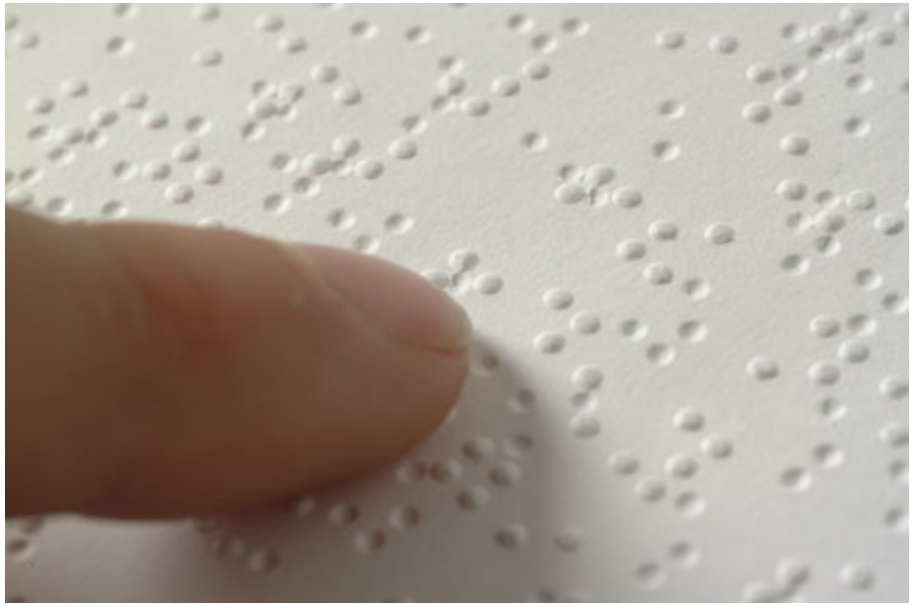
Q: Is flag semaphore an artificial language?

A: No, because, like Morse code, it is just a code for the alphabet and numerals.

Clarify the Distinction

Q: Is Braille an artificial language?

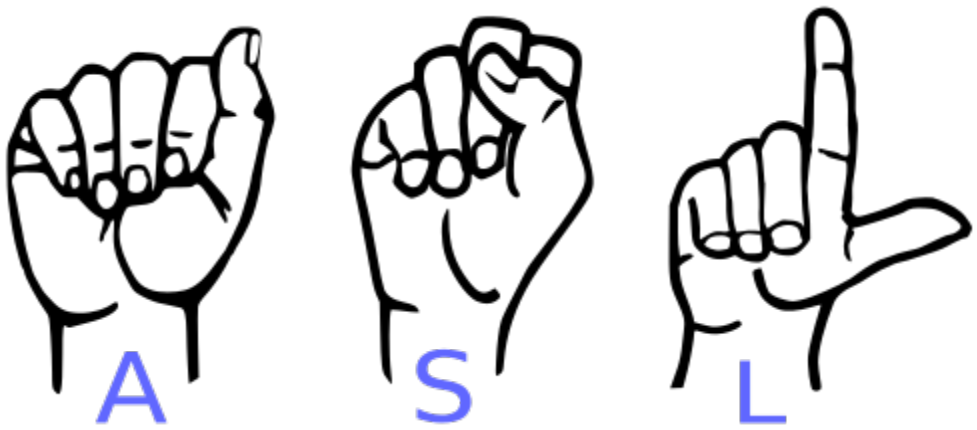
A: No, because, like Morse code, it is just a code for the alphabet and numerals.



Clarify the Distinction

Q: Is ASL (American Sign Language) an artificial language?

A: No. Even though much of it was constructed, the core of ASL traces back to “Old French Sign Language,” which had developed naturally among a community of about 200 deaf persons in Paris by the 18th century.



Why Do Us CompSci People *Love* Artificial Languages?

Because they are usually designed to be:

- Concise
- Readily parsed
- Unambiguous
- Subject to regular, consistent rules of interpretation

Why Do *Natural Languages* Drive Us CompSci People Crazy Sometimes?

Because we have to deal with:

- Large vocabulary
- Complex syntax
- Irregularities
- Ambiguous semantics
- Even more problems such as humor, irony, metaphor, connotation, neologisms

What This Means

The foregoing problems in natural language can be taken to indicate that the very notion of “NLP” is almost an oxymoron:

> NLP is “natural language processing (by machine),” but natural languages were *not designed to be processed by machine!*

And yet, it can be done.

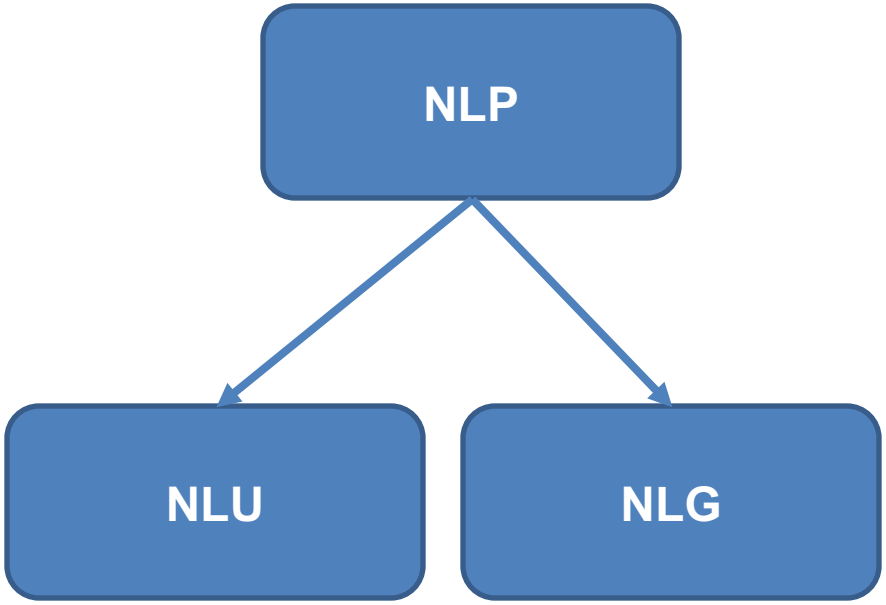
DataScience@SMU

Two Sides of NLP

Introduction to Natural Language Processing

Two Sides of NLP

The “P” in “NLP” just means processing, but there are different kinds of processing.



Two Sides of NLP

NLU: natural language understanding

We try to get the machine to understand some inputted natural language.

Usually, NLU is demonstrated by production of a structured interpretation of the input.

Wait…

Strictly speaking, does NLU really exist?

Could a machine really “understand” natural language? That’s a philosophical question that gives even geniuses like Leibniz here a hard time.

Wait…

Strictly speaking, does NLU really exist?

Regardless of the answer, we can make machines output a representation of inputted text, where that representation indicates to us humans something meaningful about the text. So the understanding may be always *ours*, not *the machine's.* And that's OK.

Two Sides of NLP

So let's go back and revise our working definition of
NLU: natural language understanding.

We try to get the machine to ***produce a useful representation*** of some inputted natural language.

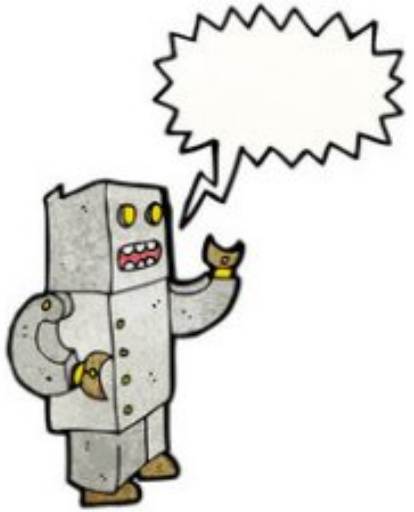
That will do just fine.

Two Sides of NLP

The other side of NLP is
NLG: natural language generation.

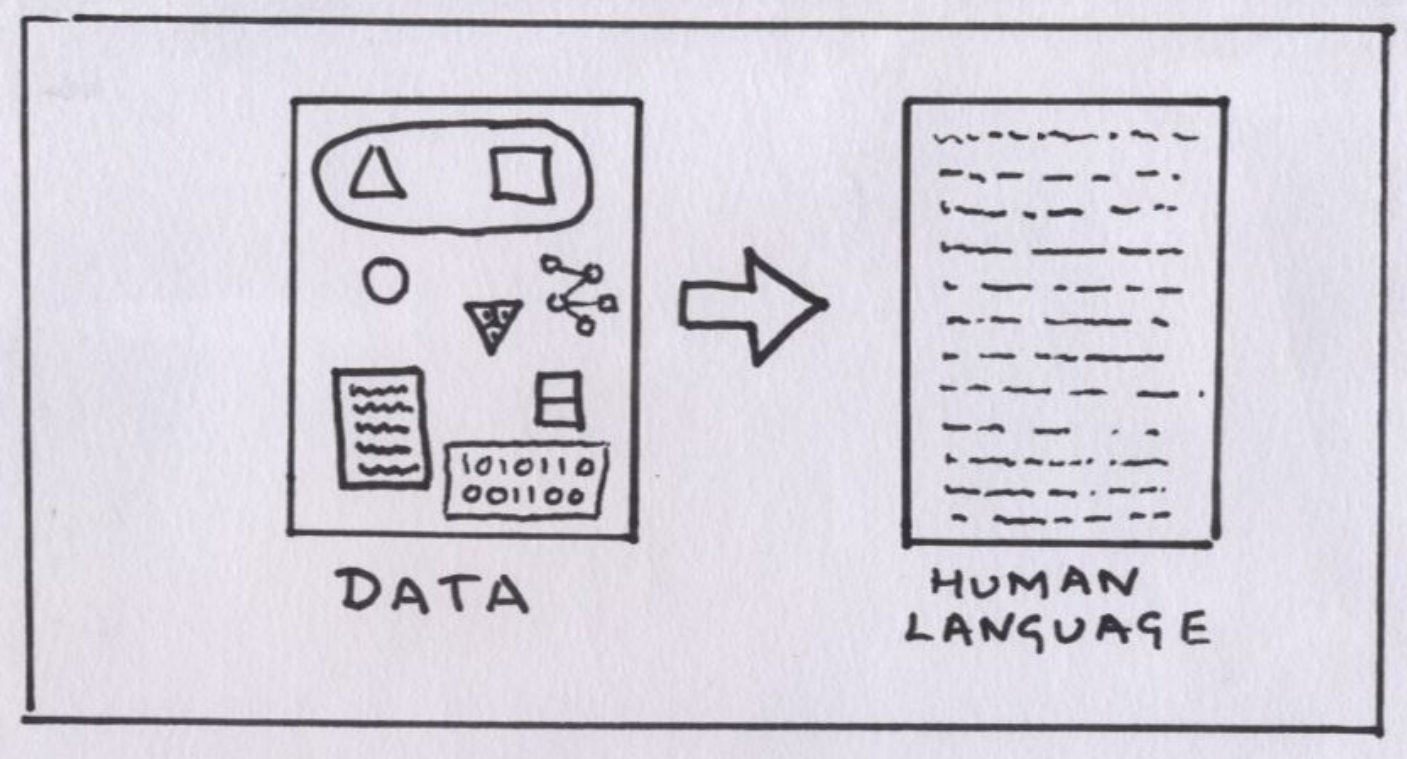
We try to get the machine to produce usable, natural language output that is not just identical to its input.

Similarly, we can put aside the question of whether the machine "understands" or "intends" the content it produces by way of NLG.



Two Sides of NLP

With NLG we create natural-seeming language from structured data.



Observation:
Rhetoric in the Industry Today

- You will sometimes hear people throw around the acronym NLP as though it means just NLU. Why might that be?
- Perhaps because, in data science we do utilize NLU a lot more than we do NLG.
  - This is because our chief use case of NLP, as data scientists, is to glean manipulable data from a pile of unstructured text. And it is NLU, not NLG, that does that for us.
- So because of our preoccupation with NLU, we sometimes forget about NLG and use the terms NLP and NLU interchangeably.
  - That's OK, but we must remember that NLP on the whole encompasses both NLU and NLG.

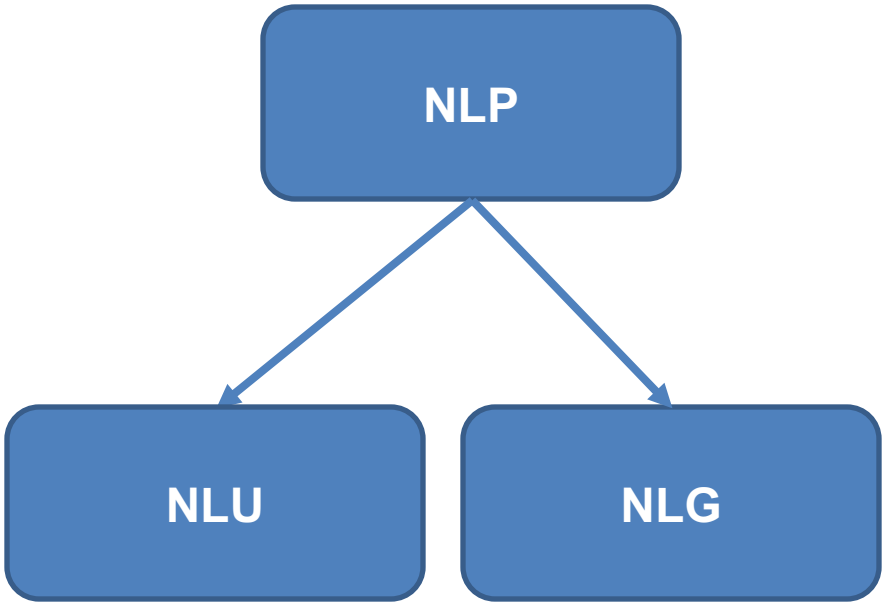
DataScience@SMU

Applications of NLU

Introduction to Natural Language Processing

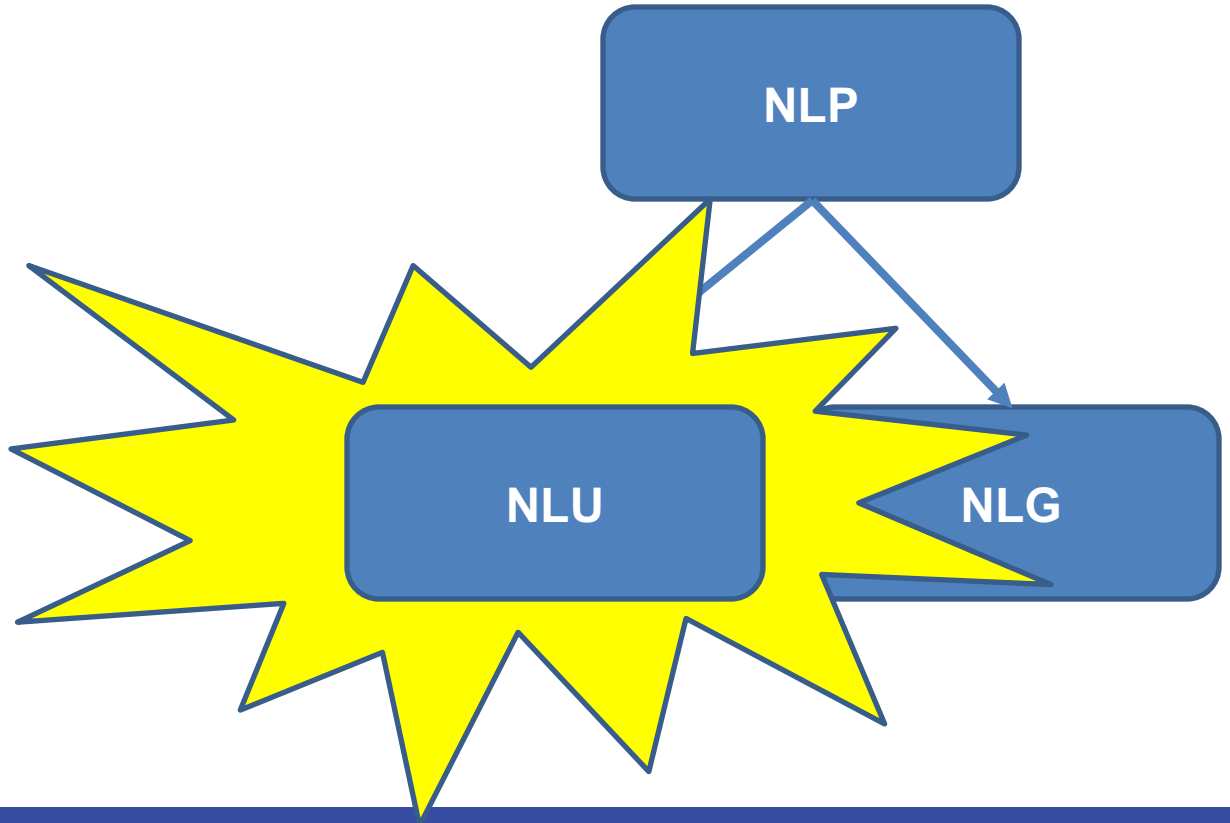
Two Sides of NLP

Let's look at applications of the first side of NLP, which is NLU:
natural language understanding.



Two Sides of NLP

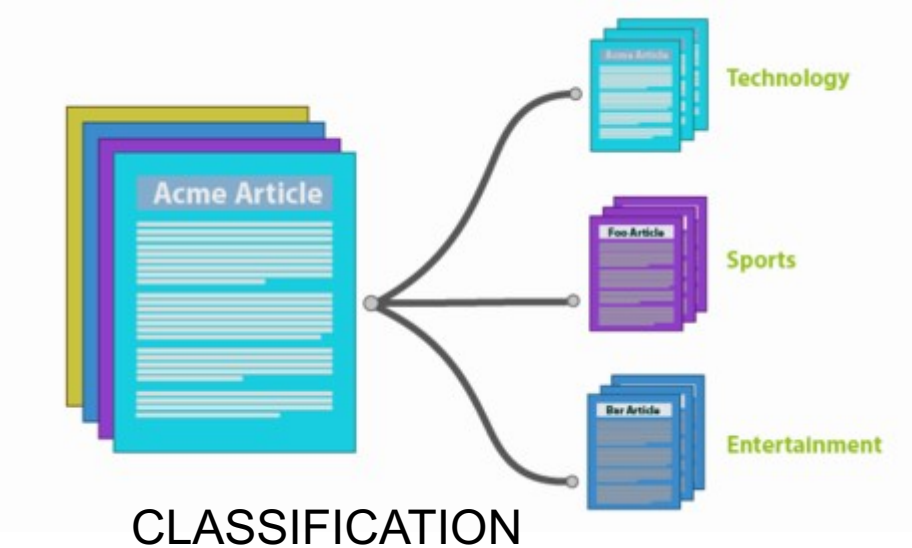
Let's look at applications of the first side of NLP, which is NLU:
natural language understanding.



Applications of NLU

Text annotation

- Tagging
- Metadata extraction/generation
- Classification
- Document summarization

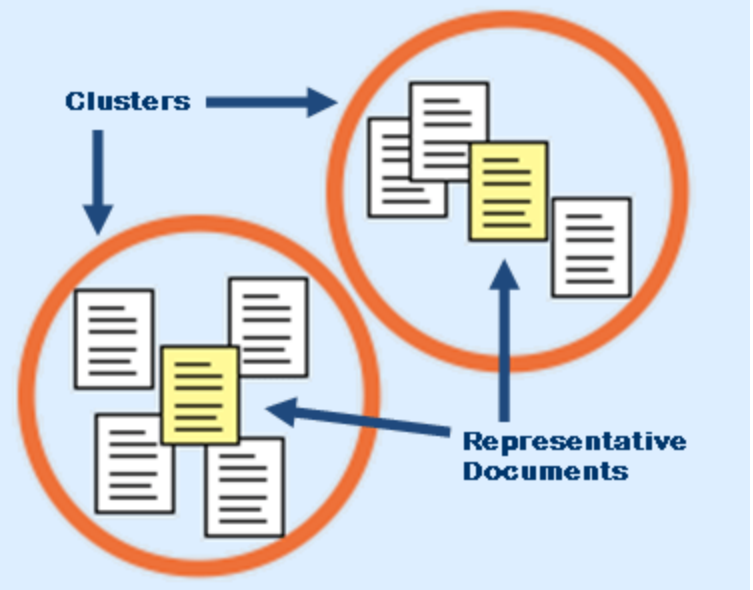


CLASSIFICATION

Applications of NLU

Corpus analytics

- Theme extraction
- Clustering
- Taxonomy mapping
- Sentiment analysis



CLUSTERING

Applications of NLU

Search applications

- Query repair
- Query refinement
- Results postprocessing, e.g., ranking, clustering, encapsulation (as below)

GOOGLE:

Apple Pie Recipe | Taste of Home

https://www.tasteofhome.com/recipes/apple-pie

★★★★½ Rating: 4.5 - 151 votes - 1 hr 5 min - 414 cal

Ingredients. 1/2 cup sugar. 1/2 cup packed brown sugar. 3 tablespoons all-purpose flour. 1 teaspoon ground cinnamon. 1/4 teaspoon ground ginger. 1/4 teaspoon ground nutmeg. 6 to 7 cups thinly sliced peeled tart **apples**. 1 tablespoon lemon juice.

vs.

BING:

Apple Pie Recipe | Taste of Home

www.tasteofhome.com › Recipes › NEXT RECIPE

I remember coming home sullen one day because we'd lost a softball game. Grandma, in her wisdom, suggested, "Maybe a slice of my homemade **apple pie** will make you …

5/5 ★★★★★ (129) **Author:** Taste of Home

Applications of NLU

Advanced applications

- Machine translation
- Knowledge discovery
- Question handling
  - Question typing/routing
  - Question-FAQ approximate matching

FAQ

1. **How do I…**
2. **Where is the…**
3. **What is the…**
4. **When do I…**
5. **Do you have a…**

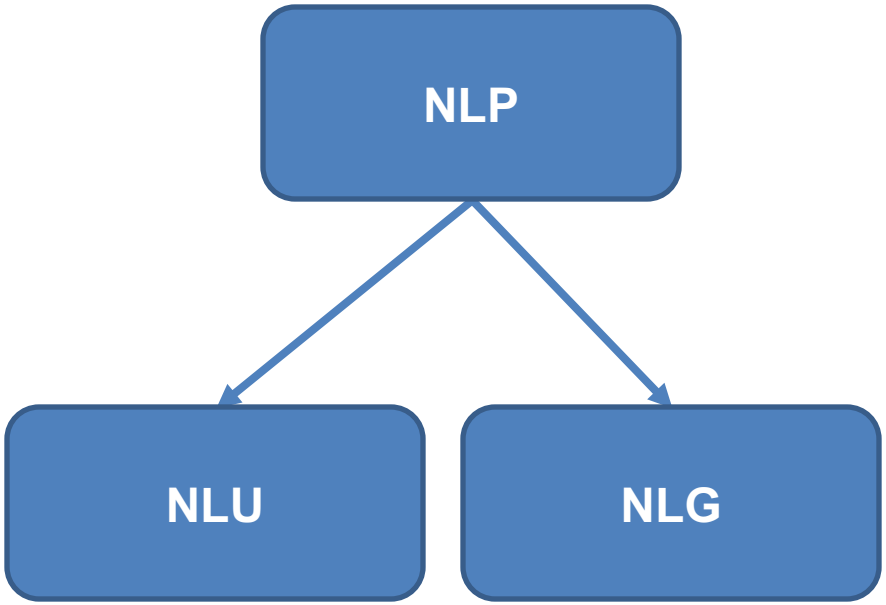
DataScience@SMU

Applications of NLG

Introduction to Natural Language Processing

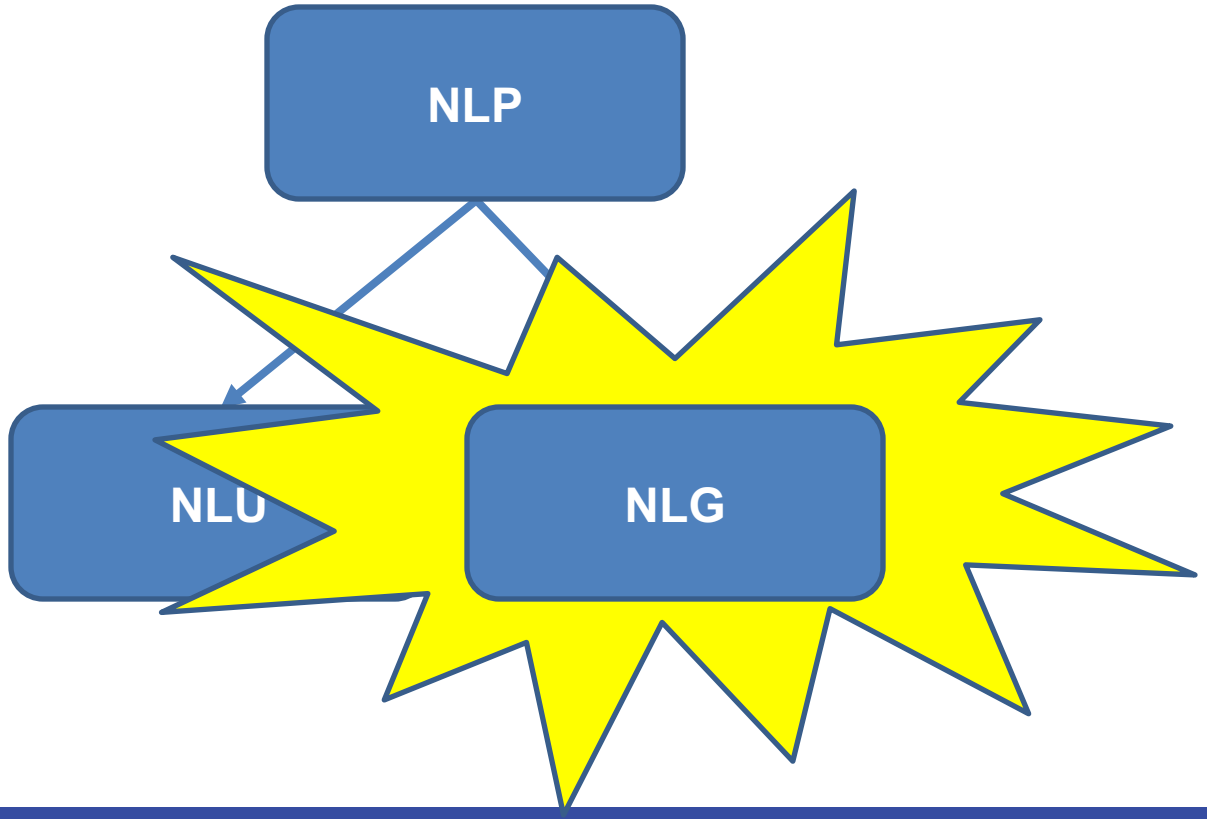
Two Sides of NLP

Now let's look at applications of the second side of NLP, which is NLG: natural language generation.



Two Sides of NLP

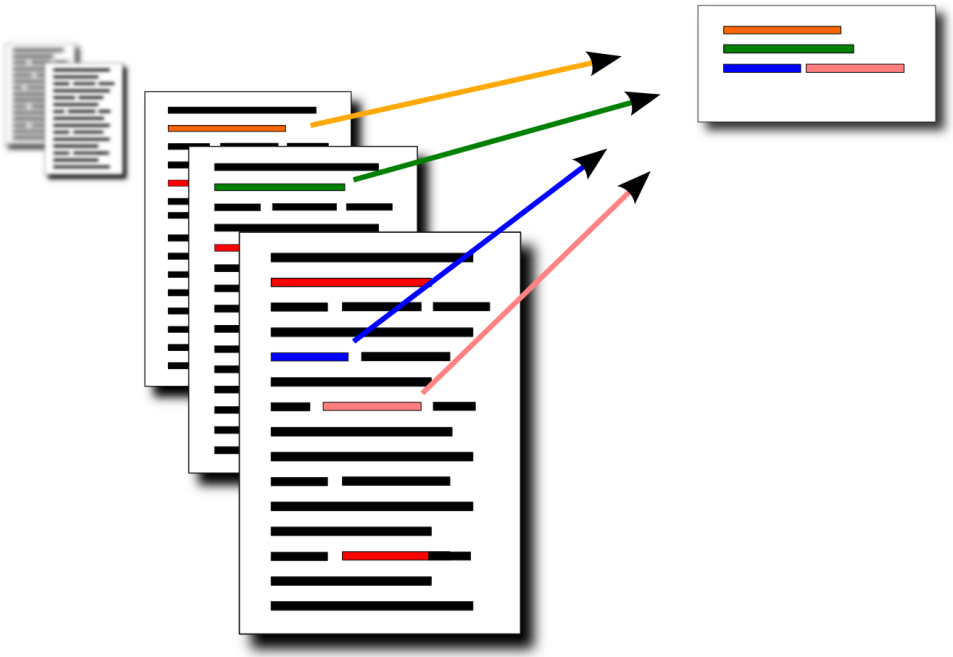
Now let's look at applications of the second side of NLP, which is NLG: natural language generation.



Applications of NLG

Text annotation

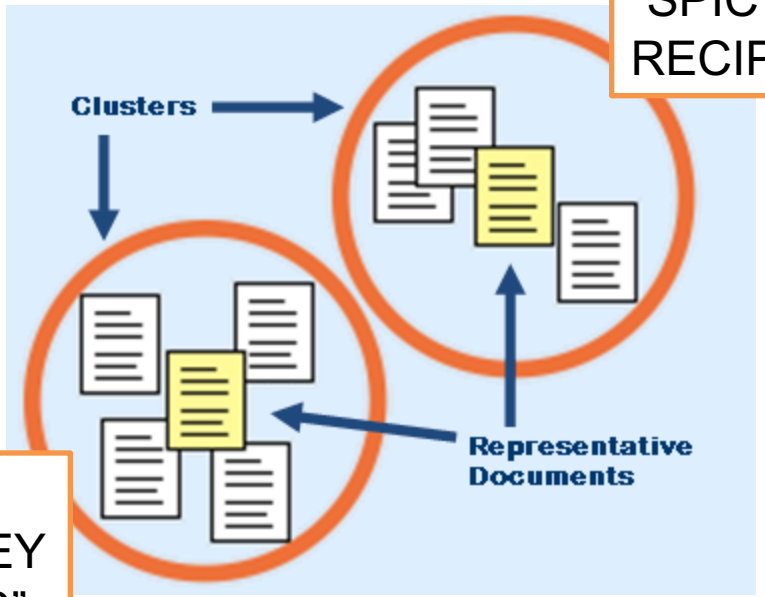
- Document summarization
- Generation of callouts/headlines



Applications of NLG

Corpus analytics

- Labelling of clusters
- Synopsizing of corpus-wide topic and/or sentiment trends



Applications of NLG

Search applications

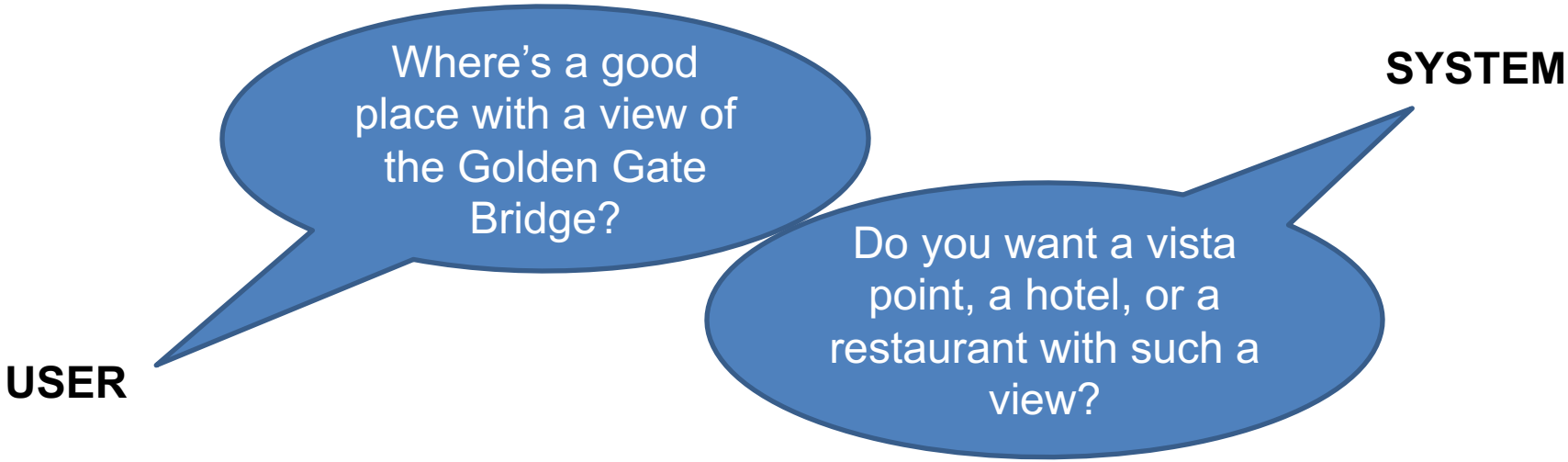
- Advanced capsule generation
  (summarization modified to fit the query)
- Advanced query refinement
  (next-gen version of “did you mean?” for disambiguation of query’s meaning)

QUERY: “review of the hobbit”
RESPONSE: *Do you mean the movie or the book?*

Applications of NLG

Advanced applications

- Machine translation
  (construction in target language)
- Knowledge discovery
  (human-friendly presentation)
- Question handling
  - Question refinement
  - Question answering



DataScience@SMU