

Answering Questions

Grzegorz Chrupała

Language in the Digital Age

What is the capital of Syria?

What is the capital of Syria?



Damascus

Who discovered penicillin?

Who discovered penicillin?

Alexander Fleming

Biologist

Sir Alexander Fleming, FRSE, FRS, FRCS was a Scottish biologist, pharmacologist and botanist. He wrote many articles on bacteriology, immunology, and chemotherapy. [Wikipedia](#)



Born: August 6, 1881, [Lochfield](#)

Died: March 11, 1955, [London, United Kingdom](#)

Awards: Nobel Prize in Physiology or Medicine, John Scott Legacy Medal and Premium

Who is Angela Merkel married to?

Who is Angela Merkel married to?

Angela Merkel Spouse



Joachim Sauer
(m. 1998)

What is the meaning of life?

What is the meaning of life?

42

Who discovered penicillin?

Who discovered penicillin?

What if we just look for passages which mention **penicillin**.

- **Penicillin** (sometimes abbreviated PCN or pen) is a group of antibiotics derived from *Penicillium* fungi.

- **Penicillin** (sometimes abbreviated PCN or pen) is a group of antibiotics derived from Penicillium fungi.
- **Penicillin** is one of the first and still one of the most widely used antibiotic agents - derived from the Penicillium mold.

- **Penicillin** (sometimes abbreviated PCN or pen) is a group of antibiotics derived from Penicillium fungi.
- **Penicillin** is one of the first and still one of the most widely used antibiotic agents - derived from the Penicillium mold.
- **Penicillin** is used to treat infections caused by bacteria, such as ear infections.

- **Penicillin** (sometimes abbreviated PCN or pen) is a group of antibiotics derived from *Penicillium* fungi.
- **Penicillin** is one of the first and still one of the most widely used antibiotic agents - derived from the *Penicillium* mold.
- **Penicillin** is used to treat infections caused by bacteria, such as ear infections.
- **Penicillins** are a certain collection of antibiotics that eliminate infection causing bacteria.

Not relevant

What if we look for
passages which
mention **penicillin**
and **discover**.

- Fleming was the first to **discover** the properties of the active substance, giving him the privilege of naming it: **penicillin**

- Fleming was the first to **discover** the properties of the active substance, giving him the privilege of naming it: **penicillin**
- Howard Florey did not **discover penicillin**.

- Fleming was the first to **discover** the properties of the active substance, giving him the privilege of naming it: **penicillin**
- Howard Florey did not **discover penicillin**.
- It could be argued here that Fleming did accidentally **discover penicillin** since he wasn't looking for it at that moment in time.

- Fleming was the first to **discover** the properties of the active substance, giving him the privilege of naming it: **penicillin**
- Howard Florey did not **discover penicillin**.
- It could be argued here that Fleming did accidentally **discover penicillin** since he wasn't looking for it at that moment in time.
- Did Ian Fleming **discover penicillin**?

We're getting
somewhere.

We're getting somewhere.

Can we already answer the
question?

KISS approach

Recognize mentions of PEOPLE.

KISS approach

Recognize mentions of PEOPLE.

Count how many times each mention appears.

KISS approach

Recognize mentions of PEOPLE.

Count how many times each mention appears.

Return most frequent.

Or we could try to understand
the question even better.

Or we could try to understand
the question even better.

How?

Who discovered penicillin?

Expected answer type

Expected answer type

Who discovered penicillin?
PERSON

Who/what the question is about

Who discovered penicillin?
SUBSTANCE

Relation expressed in the question

Who discovered penicillin?
X-DISCOVER-Y

Who discovered penicillin?
PERSON X-DISCOVER-Y SUBSTANCE

Who discovered penicillin?

Who discovered penicillin?

Retrieve passages which

Who discovered penicillin?

Retrieve passages which

- Talk about the **SUBSTANCE** penicillin

Who discovered penicillin?

Retrieve passages which

- Talk about the **SUBSTANCE** penicillin
- Express relation **X-DISCOVER-Y**

Who discovered penicillin?

Retrieve passages which

- Talk about the **SUBSTANCE** penicillin
- Express relation **X-DISCOVER-Y**
- Mention a **PERSON**

- The laboratory in which **Fleming** discovered and tested **penicillin** is preserved as the **Alexander Fleming** Laboratory Museum in St. Mary's Hospital, Paddington.

How about these.

- In contrast, Fleming isolated the substance penicillin from the mould *Penicillium notatum*.
- The discovery of penicillin is attributed to Scottish scientist and Nobel laureate Alexander Fleming in 1928.
- Duchesne cured typhoid, but the penicillin isolated by Fleming does not cure typhoid.

Variations

- In contrast, **Fleming isolated** the substance **penicillin** from the mould *Penicillium notatum*.
- The **discovery of penicillin** is attributed to **Scottish scientist and Nobel laureate Alexander Fleming** in 1928.
- **Duchesne** cured typhoid, but **the penicillin isolated by Fleming** does not cure typhoid.

Lexical variation

Morphology

discover discovering discovered
discoverer discovery rediscover

penicillin penicillins

Morphology

discover discovering discovered
discoverer discovery rediscover

penicillin penicillins

Synonymy

discover find
isolate invent develop originate

penicillin PCN pen
Penicilium mold agent

Syntactic variation

Syntactic variation

X discovered Y

Syntactic variation

X discovered Y
Y was discovered by X

Syntactic variation

X discovered Y
Y was discovered by X
discovery of Y by X

Syntactic variation

X discovered Y
Y was discovered by X
discovery of Y by X
X, who discovered Y, ...

Syntactic variation

X discovered Y
Y was discovered by X
discovery of Y by X
X, who discovered Y, ...
Y, which X discovered, ...

Syntactic variation

X discovered Y

Y was discovered by X

discovery of Y by X

X, who discovered Y, ...

Y, which X discovered, ...

Y, which was discovered by X, ...

Syntactic variation

X discovered Y

Y was discovered by X

discovery of Y by X

X, who discovered Y, ...

Y, which X discovered, ...

Y, which was discovered by X, ...

X was the discovered of Y

Syntactic variation

X discovered Y

Y was discovered by X

discovery of Y by X

X, who discovered Y, ...

Y, which X discovered, ...

Y, which was discovered by X, ...

X was the discovered of Y

X, the discoverer of Y, ...

Syntactic variation

X discovered Y

Y was discovered by X

discovery of Y by X

X, who discovered Y, ...

Y, which X discovered, ...

Y, which was discovered by X, ...

X was the discovered of Y

X, the discoverer of Y, ...

Y discoverer X

X isolated Y
Y was found by X
isolation of Y by X
X, who invented Y, ...
Y, which X found, ...
Y, which was originated by X, ...
X was the isolated of Y
X, the inventor of Y, ...
Y inventor X

We need **generalization**

How can we generalize over
lexical variation?

Discover words with similar meanings

Discover words with similar meanings

Words with similar meanings occur in similar contexts

Distributional semantics

Penicillin

antibiotics

causes

symptoms

fungi

allergies

wonder

fungal

Fleming

Penicillin

antibiotics

causes

symptoms

fungi

allergies

wonder

fungal

Fleming

PCN

antibiotics

fungi

allergies

dosage

lactamase

drugs

bacteria

medication

Penicillin PCN

antibiotics	10	15
fungi	8	3
causes	7	1
allergies	7	7
symptoms	5	1
fungal	3	1
bacteria	3	5
...
astronaut	0.01	0.0

Word classes

Word classes

Clustering

Flat clustering

Flat clustering

Berlin Bangkok Tokyo Warsaw

Flat clustering

Berlin Bangkok Tokyo Warsaw

Sarkozy Merkel Obama Berlusconi

Flat clustering

Berlin Bangkok Tokyo Warsaw

Sarkozy Merkel Obama Berlusconi

Mr Ms President Dr

Agglomerative clustering algorithm

Agglomerative clustering algorithm

Given a way to measure similarity between clusters of words:

Agglomerative clustering algorithm

Given a way to measure similarity between clusters of words:

- 1 Put each word into a separate cluster.

Agglomerative clustering algorithm

Given a way to measure similarity between clusters of words:

- 1 Put each word into a separate cluster.
- 2 Look for two most similar clusters and merge them.

Agglomerative clustering algorithm

Given a way to measure similarity between clusters of words:

- 1 Put each word into a separate cluster.
- 2 Look for two most similar clusters and merge them.
- 3 If there is only one cluster, stop.

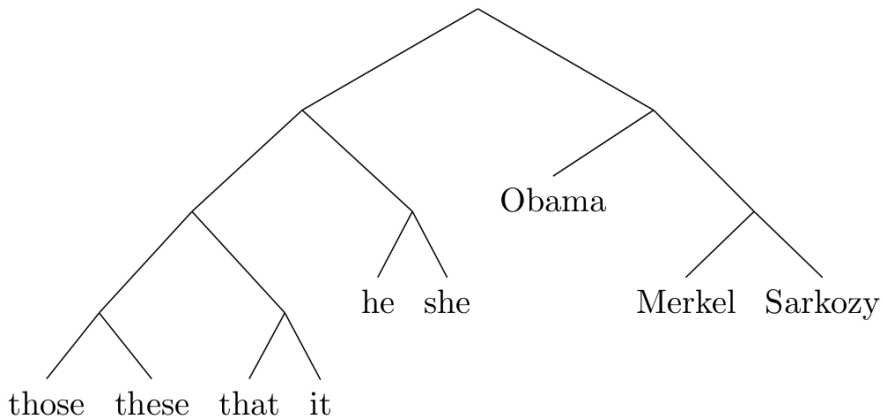
Agglomerative clustering algorithm

Given a way to measure similarity between clusters of words:

- 1 Put each word into a separate cluster.
- 2 Look for two most similar clusters and merge them.
- 3 If there is only one cluster, stop.
Othewise, go to step 2.

those these that it he she Obama Merkel Sarkozy

A possible result



After clustering, **penicillin** and **PCN** will hopefully end up in the same cluster.

After clustering, **penicillin** and **PCN** will hopefully end up in the same cluster.

Retrieve passages containing

After clustering, **penicillin** and **PCN** will hopefully end up in the same cluster.

Retrieve passages containing
{discover, find, invent, isolate, ...}

After clustering, **penicillin** and **PCN** will hopefully end up in the same cluster.

Retrieve passages containing
{discover, find, invent, isolate, ...}
and

After clustering, **penicillin** and **PCN** will hopefully end up in the same cluster.

Retrieve passages containing
{discover, find, invent, isolate, ...}
and
{PCN, penicillin, Penicilium mold agent,
Beta-lactamase inhibitor, ...}

After clustering, **penicillin** and **PCN** will hopefully end up in the same cluster.

Retrieve passages containing
{discover, find, invent, isolate, ...}
and
{PCN, penicillin, Penicilium mold agent,
Beta-lactamase inhibitor, ...}
and mentioning a **PERSON**.

Fleming discovered penicillin but Duchesne cured typhoid.

Do we want Fleming or Duchesne?

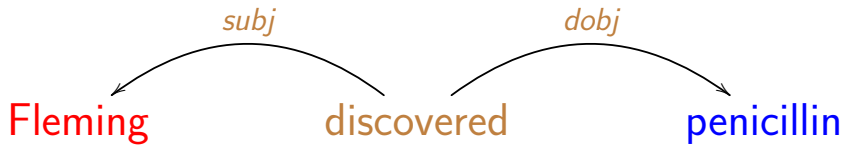
Answer extraction

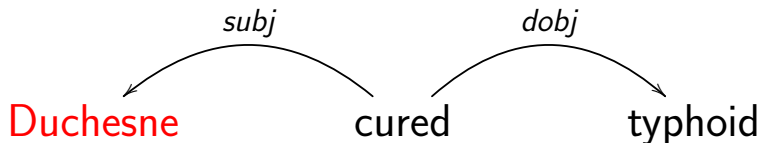
Syntactic analysis

Syntactic analysis aka **Parsing**

Look for the **PERSON** who is a **subject** of the verb **discover** whose **object** is **penicillin**.







Do we need to worry about syntactic variability?

Do we need to worry about syntactic variability?

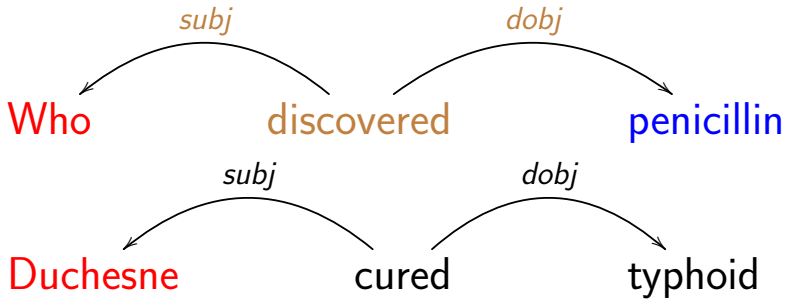
To some degree

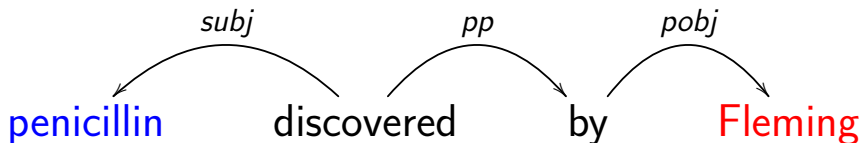
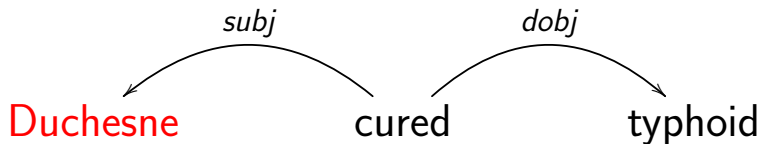
Duchesne cured typhoid, but the penicillin discovered by Fleming does not cure typhoid.

Duchesne cured typhoid, but **the penicillin** discovered by **Fleming** does not cure typhoid.

How do we choose between **Duchesne** and **Fleming**?







Solution?

We need **generalization** over
syntactic constructions.

We could apply clustering techniques similar to those for clustering words.

	X discovered Y	Y discovered by X
X=scientist	15	14
X=Columbus	10	11
Y=America	9	10
Y=antibiotic	8	8
X=Tombaugh	4	5
Y=Pluto	3	2
Y=radium	1	1
X=Curie	1	1
...

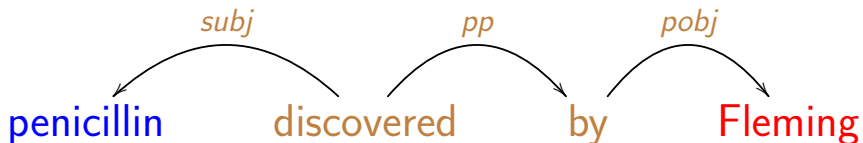
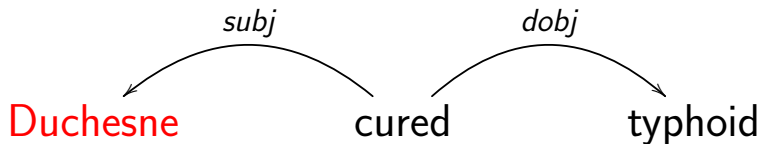


Figure out **Expected Answer Type**.

Figure out **Expected Answer Type**.

Identify **entities** of this type in answers.

Figure out **Expected Answer Type**.

Identify **entities** of this type in answers.

Identify **relations and slots** while abstracting from lexical and syntactic variability.

Figure out **Expected Answer Type**.

Identify **entities** of this type in answers.

Identify **relations and slots** while abstracting from lexical and syntactic variability.

Extract answer by **matching** question **structure** to passage.

Playful Jopardy questions

Playful Jopardy questions

“A long winded argument delivered by a frothy pie topping.”

Playful Jopardy questions

“A long winded argument delivered by a frothy pie topping.”

“Meringue Harangue.”

Finding opinions

Finding opinions

What do people think about iPhone.

Finding opinions

What do people think about iPhone.

- iPhone is getting faster year after year!
- Bigger all the iPhone-wielding Apple fanbois...
- i soooo want the new gold iphone!!