

Bachelor`s thesis defense

Semi-supervised Cleansing of Web-based Argument Corpora

Jonas Dorsch

September 09, 2020



Contents

- Introduction
- Approach
- Evaluation
- Conclusion

Introduction

Google

abortion

× |  

[All](#) [Images](#) [News](#) [Videos](#) [Maps](#) [More](#) [Settings](#) [Tools](#)

About 159,000,000 results (0,59 seconds)

[en.wikipedia.org](#) > [wiki](#) > [Abortion](#) ▾
Abortion - Wikipedia
Abortion is the ending of a pregnancy by removal or expulsion of an embryo or fetus. An abortion that occurs without intervention is known as a miscarriage or ...
Other names: Induced miscarriage, termination ...
[History of abortion](#) · [Abortion law](#) · [Abortion debate](#) · [Medical abortion](#)
You've visited this page 2 times. Last visit: 4/8/19

[www.nhs.uk](#) > [conditions](#) > [abortion](#) ▾
Abortion - NHS
How to get an abortion. Abortions can only be carried out under the care of an NHS hospital or a licensed clinic, and are usually available free of charge on the ...


People also ask

What states are abortions illegal in?	▼
What is abortion and miscarriage?	▼
What is abortion simple?	▼
How many weeks can you have an abortion in Louisiana?	▼


[Feedback](#)

Top stories


[Federal judge blocks rollout of Tennessee's strict new abortion restrictions as court weighs law's fate](#)
The Tennessean · 2 days ago



[The Ghost of Margaret Sanger](#)
The New York Times · 1 day ago



[Girl, 13, had 'abortion' chanted at her after she terminated pregnancy from brutal rape](#)
The Mirror · 6 hours ago



→ [More for abortion](#)

Introduction

Google

abortion

× | 🔊 🔍

[All](#) [Images](#) [News](#) [Videos](#) [Maps](#) [More](#) [Settings](#) [Tools](#)

About 159,000,000 results (0,59 seconds)

[en.wikipedia.org](#) > [wiki](#) > [Abortion](#) ▾
Abortion - Wikipedia
Abortion is the ending of a pregnancy by removal or expulsion of an embryo or fetus. An abortion that occurs without intervention is known as a miscarriage or ...
Other names: Induced miscarriage, termination ...
[History of abortion](#) · [Abortion law](#) · [Abortion debate](#) · [Medical abortion](#)
You've visited this page 2 times. Last visit: 4/8/19

[www.nhs.uk](#) > [conditions](#) > [abortion](#) ▾
Abortion - NHS
How to get an abortion. Abortions can only be carried out under the care of an NHS hospital or a licensed clinic, and are usually available free of charge on the ...

People also ask

What states are abortions illegal in? ▾

What is abortion and miscarriage? ▾

What is abortion simple? ▾

How many weeks can you have an abortion in Louisiana? ▾

[Feedback](#)

Top stories

[Federal judge blocks rollout of Tennessee's strict new abortion restrictions as court weighs law's fate](#)
The Tennessean · 2 days ago

[The Ghost of Margaret Sanger](#)
The New York Times · 1 day ago

[Girl, 13, had 'abortion' chanted at her after she terminated pregnancy from brutal rape](#)
The Mirror · 6 hours ago

→ [More for abortion](#)

Definitions

Introduction

Google

abortion

× | 🔊 🔍

🔍 All 🖼️ Images 📰 News 📺 Videos 📍 Maps ⋮ More Settings Tools

About 159,000,000 results (0,59 seconds)

en.wikipedia.org › wiki › Abortion ▾
Abortion - Wikipedia
Abortion is the ending of a pregnancy by removal or expulsion of an embryo or fetus. An abortion that occurs without intervention is known as a miscarriage or ...
Other names: Induced miscarriage, termination ...
[History of abortion](#) · [Abortion law](#) · [Abortion debate](#) · [Medical abortion](#)
You've visited this page 2 times. Last visit: 4/8/19

www.nhs.uk › conditions › abortion ▾
Abortion - NHS
How to get an abortion. Abortions can only be carried out under the care of an NHS hospital or a licensed clinic, and are usually available free of charge on the ...


People also ask

What states are abortions illegal in?	▼
What is abortion and miscarriage?	▼
What is abortion simple?	▼
How many weeks can you have an abortion in Louisiana?	▼


Feedback

Top stories


[Federal judge blocks rollout of Tennessee's strict new abortion restrictions as court weighs law's fate](#)
The Tennessean · 2 days ago



[The Ghost of Margaret Sanger](#)
The New York Times · 1 day ago



[Girl, 13, had 'abortion' chanted at her after she terminated pregnancy from brutal rape](#)
The Mirror · 6 hours ago



→ [More for abortion](#)

Definitions

Related questions

Introduction

Google

abortion

× | 🔊 🔍

🔍 All 🖼️ Images 📰 News 📺 Videos 📍 Maps ⋮ More Settings Tools

About 159,000,000 results (0,59 seconds)

en.wikipedia.org › wiki › Abortion ▾
Abortion - Wikipedia
Abortion is the ending of a pregnancy by removal or expulsion of an embryo or fetus. An abortion that occurs without intervention is known as a miscarriage or ...
Other names: Induced miscarriage, termination ...
[History of abortion](#) · [Abortion law](#) · [Abortion debate](#) · [Medical abortion](#)
You've visited this page 2 times. Last visit: 4/8/19

www.nhs.uk › conditions › abortion ▾
Abortion - NHS
How to get an abortion. Abortions can only be carried out under the care of an NHS hospital or a licensed clinic, and are usually available free of charge on the ...


People also ask

What states are abortions illegal in?	▼
What is abortion and miscarriage?	▼
What is abortion simple?	▼
How many weeks can you have an abortion in Louisiana?	▼


Feedback

Top stories


[Federal judge blocks rollout of Tennessee's strict new abortion restrictions as court weighs law's fate](#)
The Tennessean · 2 days ago



[The Ghost of Margaret Sanger](#)
The New York Times · 1 day ago



[Girl, 13, had 'abortion' chanted at her after she terminated pregnancy from brutal rape](#)
The Mirror · 6 hours ago



→ [More for abortion](#)

Definitions

Related questions

News

Introduction

Google

Good at:

Presenting factual data

Bad at:

Presenting argumenative data

The screenshot shows a Google search for 'abortion'. The search bar at the top contains the word 'abortion'. Below the search bar, there are tabs for 'All', 'Images', 'News', 'Videos', 'Maps', and 'More'. The search results are categorized into three main sections, each highlighted with an orange border and a label to its right:

- Definitions:** This section includes two results. The first is from Wikipedia, titled 'Abortion - Wikipedia', with a brief definition: 'Abortion is the ending of a pregnancy by removal or expulsion of an embryo or fetus. An abortion that occurs without intervention is known as a miscarriage or ...'. The second result is from the NHS, titled 'Abortion - NHS', with a brief description: 'How to get an abortion. Abortions can only be carried out under the care of an NHS hospital or a licensed clinic, and are usually available free of charge on the ...'.
- Related questions:** This section is titled 'People also ask' and contains four questions with expandable answers:
 - What states are abortions illegal in?
 - What is abortion and miscarriage?
 - What is abortion simple?
 - How many weeks can you have an abortion in Louisiana?
- News:** This section is titled 'Top stories' and contains three news items, each with a headline, source, and time ago, accompanied by a small image:
 - 'Federal judge blocks rollout of Tennessee's strict new abortion restrictions as court weighs law's fate' from The Tennessean, 2 days ago.
 - 'The Ghost of Margaret Sanger' from The New York Times, 1 day ago.
 - 'Girl, 13, had 'abortion' chanted at her after she terminated pregnancy from brutal rape' from The Mirror, 6 hours ago.

At the bottom of the search results, there is a link that says '→ More for abortion'.

Do we need argumentation?

Introduction – Do we need argumentation?

- Essential part of:
 - Politics



Introduction – Do we need argumentation?

- Essential part of:
 - Politics
 - Work
 - Education



Source: <https://pixabay.com/photos/workplace-team-business-meeting-1245776/>

Introduction – Do we need argumentation?

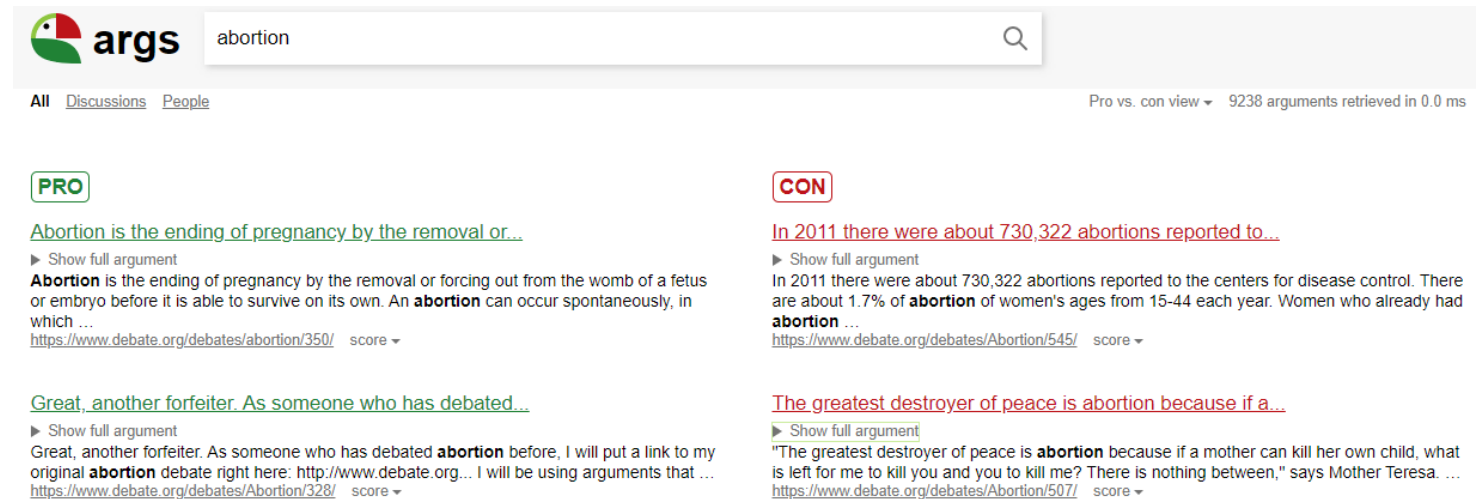
- Essential part of:
 - Politics
 - Work
 - Education
 - Relationships
 - Social life
 - etc.



Source: <https://pixabay.com/illustrations/conversation-talk-talking-people-799448/>

So why don't make arguments accessible online?

Introduction – Argument search engines



The screenshot shows the 'args' search engine interface. At the top, there is a search bar with the query 'abortion' and a magnifying glass icon. Below the search bar, there are tabs for 'All', 'Discussions', and 'People'. On the right, it says 'Pro vs. con view' and '9238 arguments retrieved in 0.0 ms'. The results are divided into two columns: 'PRO' and 'CON'. Each column shows a list of arguments with their titles, a 'Show full argument' link, and a URL. The 'PRO' column has two arguments, and the 'CON' column has two arguments.

PRO

[Abortion is the ending of pregnancy by the removal or...](#)
 ► Show full argument
Abortion is the ending of pregnancy by the removal or forcing out from the womb of a fetus or embryo before it is able to survive on its own. An **abortion** can occur spontaneously, in which ...
<https://www.debate.org/debates/abortion/350/> score ▼

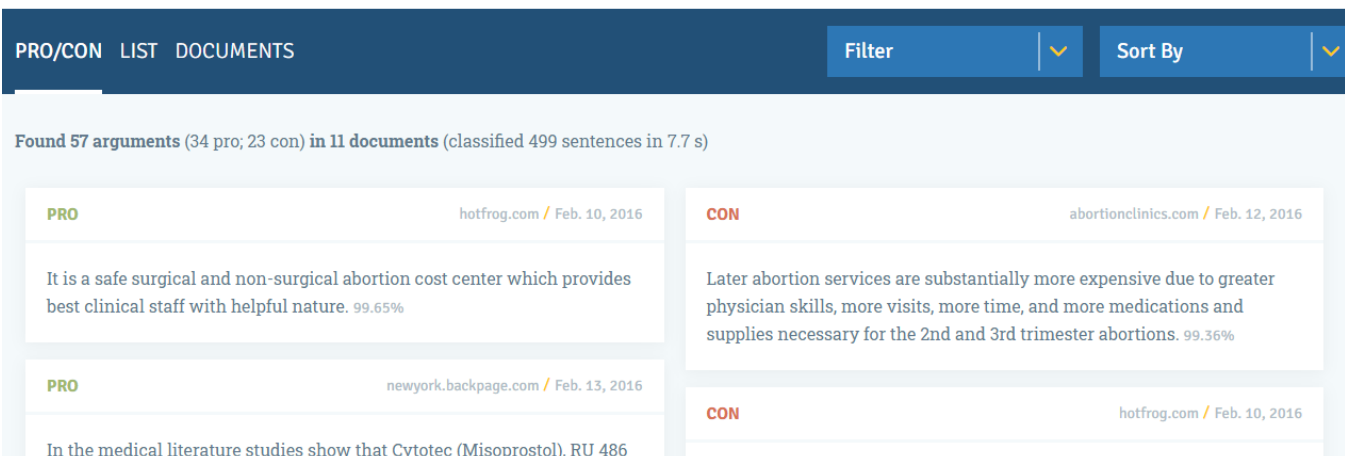
[Great, another forfeiter. As someone who has debated...](#)
 ► Show full argument
 Great, another forfeiter. As someone who has debated **abortion** before, I will put a link to my original **abortion** debate right here: <http://www.debate.org...> I will be using arguments that ...
<https://www.debate.org/debates/Abortion/328/> score ▼

CON

[In 2011 there were about 730,322 abortions reported to...](#)
 ► Show full argument
 In 2011 there were about 730,322 abortions reported to the centers for disease control. There are about 1.7% of **abortion** of women's ages from 15-44 each year. Women who already had **abortion** ...
<https://www.debate.org/debates/Abortion/545/> score ▼

[The greatest destroyer of peace is abortion because if a...](#)
 ► Show full argument
 "The greatest destroyer of peace is **abortion** because if a mother can kill her own child, what is left for me to kill you and you to kill me? There is nothing between," says Mother Teresa. ...
<https://www.debate.org/debates/Abortion/507/> score ▼

Source: <https://www.args.me/index.html>



The screenshot shows the 'ArgumentText' search engine interface. At the top, there is a navigation bar with 'PRO/CON', 'LIST', and 'DOCUMENTS'. Below this, there are buttons for 'Filter' and 'Sort By'. The main content area shows the search results for 'abortion'. It states 'Found 57 arguments (34 pro; 23 con) in 11 documents (classified 499 sentences in 7.7 s)'. The results are displayed in a table with two columns: 'PRO' and 'CON'. Each row shows the argument's title, a 'Show full argument' link, and a URL. The 'PRO' column has two arguments, and the 'CON' column has two arguments.

PRO hotfrog.com / Feb. 10, 2016

It is a safe surgical and non-surgical abortion cost center which provides best clinical staff with helpful nature. 99.65%

PRO newyork.backpage.com / Feb. 13, 2016

In the medical literature studies show that Cytotec (Misoprostol), RU 486

CON abortionclinics.com / Feb. 12, 2016

Later abortion services are substantially more expensive due to greater physician skills, more visits, more time, and more medications and supplies necessary for the 2nd and 3rd trimester abortions. 99.36%

CON hotfrog.com / Feb. 10, 2016

Source: <https://www.argumentsearch.com/>

Introduction – Argument search engines

Search Field

args abortion

All Discussions People Pro vs. con view 9238 arguments retrieved in 0.0 ms

PRO

Abortion is the ending of pregnancy by the removal or...

► Show full argument

Abortion is the ending of pregnancy by the removal or forcing out from the womb of a fetus or embryo before it is able to survive on its own. An **abortion** can occur spontaneously, in which ...
<https://www.debate.org/debates/abortion/350/> score ▼

Great, another forfeiter. As someone who has debated...

► Show full argument

Great, another forfeiter. As someone who has debated **abortion** before, I will put a link to my original **abortion** debate right here: <http://www.debate.org...> I will be using arguments that ...
<https://www.debate.org/debates/Abortion/328/> score ▼

CON

In 2011 there were about 730,322 abortions reported to...

► Show full argument

In 2011 there were about 730,322 abortions reported to the centers for disease control. There are about 1.7% of **abortion** of women's ages from 15-44 each year. Women who already had **abortion** ...
<https://www.debate.org/debates/Abortion/545/> score ▼

The greatest destroyer of peace is abortion because if a...

► Show full argument

"The greatest destroyer of peace is **abortion** because if a mother can kill her own child, what is left for me to kill you and you to kill me? There is nothing between," says Mother Teresa. ...
<https://www.debate.org/debates/Abortion/507/> score ▼

Source: <https://www.args.me/index.html>

ArgumentText EN ▼ abortion

PRO/CON LIST DOCUMENTS Filter Sort By

Found 57 arguments (34 pro; 23 con) in 11 documents (classified 499 sentences in 7.7 s)

<p>PRO hotfrog.com / Feb. 10, 2016</p> <p>It is a safe surgical and non-surgical abortion cost center which provides best clinical staff with helpful nature. 99.65%</p>	<p>CON abortionclinics.com / Feb. 12, 2016</p> <p>Later abortion services are substantially more expensive due to greater physician skills, more visits, more time, and more medications and supplies necessary for the 2nd and 3rd trimester abortions. 99.36%</p>
<p>PRO newyork.backpage.com / Feb. 13, 2016</p> <p>In the medical literature studies show that Cytotec (Misoprostol), RU 486</p>	<p>CON hotfrog.com / Feb. 10, 2016</p>

Source: <https://www.argumentsearch.com/>

Introduction – Argument search engines

Search Field

args abortion

All Discussions People Pro vs. con view 9238 arguments retrieved in 0.0 ms

PRO

Abortion is the ending of pregnancy by the removal or...

► Show full argument

Abortion is the ending of pregnancy by the removal or forcing out from the womb of a fetus or embryo before it is able to survive on its own. An **abortion** can occur spontaneously, in which ...

<https://www.debate.org/debates/abortion/350/> score ▼

Great, another forfeiter. As someone who has debated...

► Show full argument

Great, another forfeiter. As someone who has debated **abortion** before, I will put a link to my original **abortion** debate right here: <http://www.debate.org...> I will be using arguments that ...

<https://www.debate.org/debates/Abortion/328/> score ▼

CON

In 2011 there were about 730,322 abortions reported to...

► Show full argument

In 2011 there were about 730,322 abortions reported to the centers for disease control. There are about 1.7% of **abortion** of women's ages from 15-44 each year. Women who already had **abortion** ...

<https://www.debate.org/debates/Abortion/545/> score ▼

The greatest destroyer of peace is abortion because if a...

► Show full argument

"The greatest destroyer of peace is **abortion** because if a mother can kill her own child, what is left for me to kill you and you to kill me? There is nothing between," says Mother Teresa. ...

<https://www.debate.org/debates/Abortion/507/> score ▼

Source: <https://www.args.me/index.html>

Argument

ArgumentText

EN ▼ abortion

PRO/CON LIST DOCUMENTS Filter Sort By

Found 57 arguments (34 pro; 23 con) in 11 documents (classified 499 sentences in 7.7 s)

PRO hotfrog.com / Feb. 10, 2016

It is a safe surgical and non-surgical abortion cost center which provides best clinical staff with helpful nature. 99.65%

CON abortionclinics.com / Feb. 12, 2016

Later abortion services are substantially more expensive due to greater physician skills, more visits, more time, and more medications and supplies necessary for the 2nd and 3rd trimester abortions. 99.36%

PRO newyork.backpage.com / Feb. 13, 2016

In the medical literature studies show that Cytotec (Misoprostol), RU 486

CON hotfrog.com / Feb. 10, 2016

So where does the data come from?

Introduction – Argument Corpora

DEBATES ▼
OPINIONS ▼
FORUMS ▼
POLLS ▼

Sign In
Sign Up

Home > Debates > Should animals be used in labs?

Arts (1,236)

Cars (364)

Economics (2,110)

Education (6,122)

Entertainment (4,357)

Fashion (305)

Funny (1,530)

Games (1,940)

Health (3,494)

Miscellaneous (6,580)

Movies (489)

Music (851)

News (724)

People (3,153)

Philosophy (5,657)

Places-Travel (205)

Politics (16,599)

The Instigator

Con (against)

Hilal

Should animals be used in labs?

Add to My Favorites
 Report this Debate
 Share with My Friends

Do you like this debate? No Yes +0

Argument Due

We are waiting for Hilal to post argument for round #3. If you are Hilal, login to see your options.

[Hilal's Login](#)

Time Remaining

02 days **11** hours **15** minutes **56** seconds

Voting Style: Open	Point System: 7 Point
Started: 9/1/2020	Category: Science
Updated: 12 hours ago	Status: Debating Period
Viewed: 166 times	Debate No: 126163

Tweet
 Gefällt mir 0

The Contender

Pro (for)


Yoshikage69Kira


Source: <https://www.debate.org/debates/Should-animals-be-used-in-labs/1/>

Introduction – Argument Corpora

DEBATE.ORG DEBATES ▼ OPINIONS ▼ FORUMS ▼ POLLS ▼ [Sign In](#) [Sign Up](#)

[Home](#) > [Debates](#) > [Should animals be used in labs?](#)

The Instigator

Con (against)
 Hilal

The Contender

Pro (for)
 Yoshikage69Kira

Should animals be used in labs?

★ [Add to My Favorites](#) [Report this Debate](#) [Share with My Friends](#)

Do you like this debate? [No](#) [Yes](#) +0

[Hilal's Login](#)

Argument Due

We are waiting for [Hilal](#) to post argument for round #3. If you are Hilal, login to see your options.

Time Remaining
02 days 11 hours 15 minutes 56 seconds

Voting Style: Open	Point System: 7 Point
Started: 9/1/2020	Category: Science
Updated: 12 hours ago	Status: Debating Period
Viewed: 166 times	Debate No: 126163

[Tweet](#) [Gefällt mir 0](#)

Topic

Source: <https://www.debate.org/debates/Should-animals-be-used-in-labs/1/>

Introduction – Argument Corpora

The screenshot shows a debate page on DEBATE.ORG. The page title is "Should animals be used in labs?". The page is divided into three main sections: "The Instigator" (Hilal, Con (against)), "The Contender" (Yoshikage69Kira, Pro (for)), and "Argument Due". The "Argument Due" section indicates that the user Hilal is waiting to post an argument for round #3. The page also features a sidebar with various categories and a table of debate statistics.

Topic: Should animals be used in labs?

Contestants:

- The Instigator:** Hilal (Con (against))
- The Contender:** Yoshikage69Kira (Pro (for))

Argument Due: We are waiting for Hilal to post argument for round #3. If you are Hilal, login to see your options.

Time Remaining: 02 days 11 hours 15 minutes 56 seconds


Voting Style:	Open	Point System:	7 Point
Started:	9/1/2020	Category:	Science
Updated:	12 hours ago	Status:	Debating Period
Viewed:	166 times	Debate No:	126163

Annotations:

- Topic:** Points to the debate title "Should animals be used in labs?".
- Contestants:** Points to the profiles of "The Instigator" (Hilal) and "The Contender" (Yoshikage69Kira).


Introduction – Argument Corpora


Debate Rounds (3)
Comments (11)
Votes (0)



Con


It is not ethical to use animals for scientific research.


[Report this Argument](#)



Pro

Why is it not ethical? I see it as necessary for human development. If we don't use lab rats then we may never be able to develop much needed treatments for diseases.


[Report this Argument](#)



Con

Every scientific study must get consent from the participants. However, Animals are not able to give consent. Therefore, Testing on animals is a violation. Why don't we test on people who give consent? If we don't even consider this as an option, It means that we completely disregard the importance of animals lives or we think that human life is more important than animals lives.


[Report this Argument](#)



Pro

Yes, Society does place more value on humans than animals. I don't see why this would be a problem. All species place more value on members of their own kind over non members. Animals are only useful as food, Pets and for experiments, They aren't considered of the same worth as a human as they are merely cattle.


[Report this Argument](#)

This round has not been posted yet.

This round has not been posted yet.

Introduction – Argument Corpora

Debate Rounds (3) Comments (11) Votes (0)

Round 1

Con It is not ethical to use animals for scientific research. [Report this Argument](#)

Pro Why is it not ethical? I see it as necessary for human development. If we don't use lab rats then we may never be able to develop much needed treatments for diseases. [Report this Argument](#)

Round 2

Con Every scientific study must get consent from the participants. However, Animals are not able to give consent. Therefore, Testing on animals is a violation. Why don't we test on people who give consent? If we don't even consider this as an option, It means that we completely disregard the importance of animals lives or we think that human life is more important than animals lives. [Report this Argument](#)

Pro Yes, Society does place more value on humans than animals. I don't see why this would be a problem. All species place more value on members of their own kind over non members. Animals are only useful as food, Pets and for experiments, They aren't considered of the same worth as a human as they are merely cattle. [Report this Argument](#)

Round 3

This round has not been posted yet.

This round has not been posted yet.

Arguments

Introduction – The args.me corpus

- Published by Ajjour et al. [2019]
- 4 debate portals
- 390k arguments
- 60k debates

(a) Query	Absolute	Relative
climate change	251	3.5%
feminism	193	2.7%
abortion	158	2.2%
trump	146	2.0%
brexit	128	1.8%
death penalty	73	1.0%
google	58	0.8%
vegan	57	0.8%
nuclear energy	56	0.8%
donald trump	47	0.7%

(b) Conclusion	Absolute	Relative
Abortion	2 838	0.7%
Gay Marriage	1 558	0.4%
Rap Battle	1 396	0.4%
Death Penalty	972	0.3%
God exists	790	0.2%
Gun control	719	0.1%
Ivf debate	432	0.1%
Animal testing	372	0.1%
I will not contradict myself	357	0.1%
Euthanasia	331	0.1%

(a) Top 10 queries from args.me query log (from Sept 2017 to May 2019).

(b) Top 10 conclusions of arguments in the args.me corpus.

Presented in Ajjour et al. [2019]

So why do we have to cleanse the corpus?

Introduction – Corpus cleansing

Argumentative debate post pro "*abortion*"

<https://www.debate.org/debates/Abortion/526/>

Thank you for bringing up this interesting debate topic, and I look forward to an exciting debate.

This has been repeated a million times but I'm going to say it again because it must act as the foundation of this debate. Abortion is not murder, especially when it is performed before the fetus has developed into a human being. When an abortion is performed, no baby is killed or murdered. Think of it as destroying the seed before it becomes a plant. Performing an abortion is by definition not committing murder. With this fact as a base or foundation, throughout the next few rounds I will argue that abortion does not demean the value of human life, but protects it -- using facts and rational arguments.

I will also argue that mother"s, being the carrier of the fetus, have the right and responsibility to decide and weigh the advantages and disadvantages of bringing a human being into the world. Let me ask you. How is it better to have the child and allow it to lead a poor and neglected life, rather than never existing in the first place? If you value life, and the sanctity of living, then you will accept abortion.

You don"t have to practice it, advocate for it, understand it, or even respect it. But if you respect life, then you will accept abortion and allow other women to have abortions.

Thank you, I look forward to your response.

Introduction – Corpus cleansing

Argumentative debate post pro "*abortion*"

<https://www.debate.org/debates/Abortion/526/>

Irrelevant Sentences

Thank you for bringing up this interesting debate topic, and I look forward to an exciting debate.

This has been repeated a million times but I'm going to say it again because it must act as the foundation of this debate. Abortion is not murder, especially when it is performed before the fetus has developed into a human being. When an abortion is performed, no baby is killed or murdered. Think of it as destroying the seed before it becomes a plant. Performing an abortion is by definition not committing murder. With this fact as a base or foundation, throughout the next few rounds I will argue that abortion does not demean the value of human life, but protects it -- using facts and rational arguments.

I will also argue that mother"s, being the carrier of the fetus, have the right and responsibility to decide and weigh the advantages and disadvantages of bringing a human being into the world. Let me ask you. How is it better to have the child and allow it to lead a poor and neglected life, rather than never existing in the first place? If you value life, and the sanctity of living, then you will accept abortion. You don"t have to practice it, advocate for it, understand it, or even respect it. But if you respect life, then you will accept abortion and allow other women to have abortions.

Thank you, I look forward to your response.

Approach

Approach – Goal

Remove as much irrelevant sentences as possible without losing any relevant information.

Approach – Goal

Remove as much irrelevant sentences as possible without losing any relevant information.

1. High precision

Approach – Goal

Remove as much irrelevant sentences as possible without losing any relevant information.

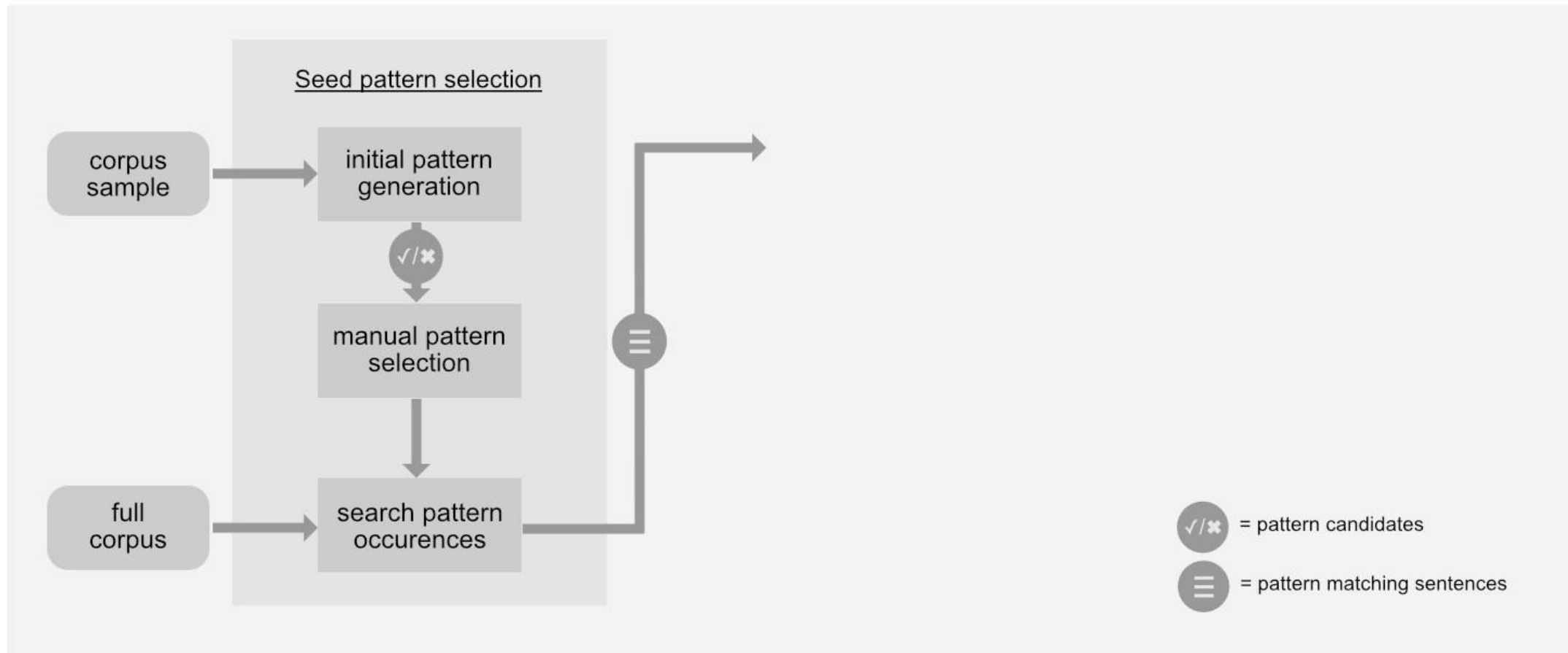
1. High precision
2. High recall

Approach – Goal

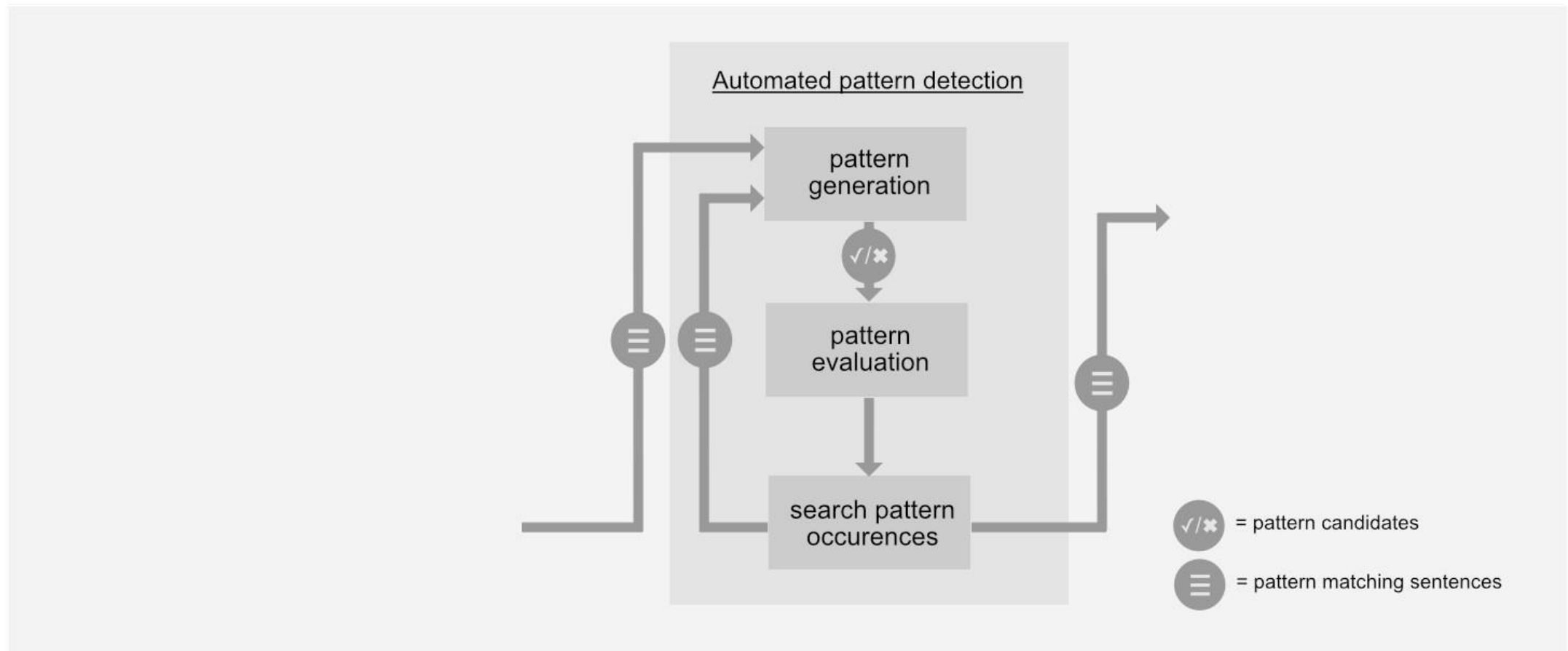
Remove as much irrelevant sentences as possible without losing any relevant information.

1. High precision
2. High recall
3. Automated

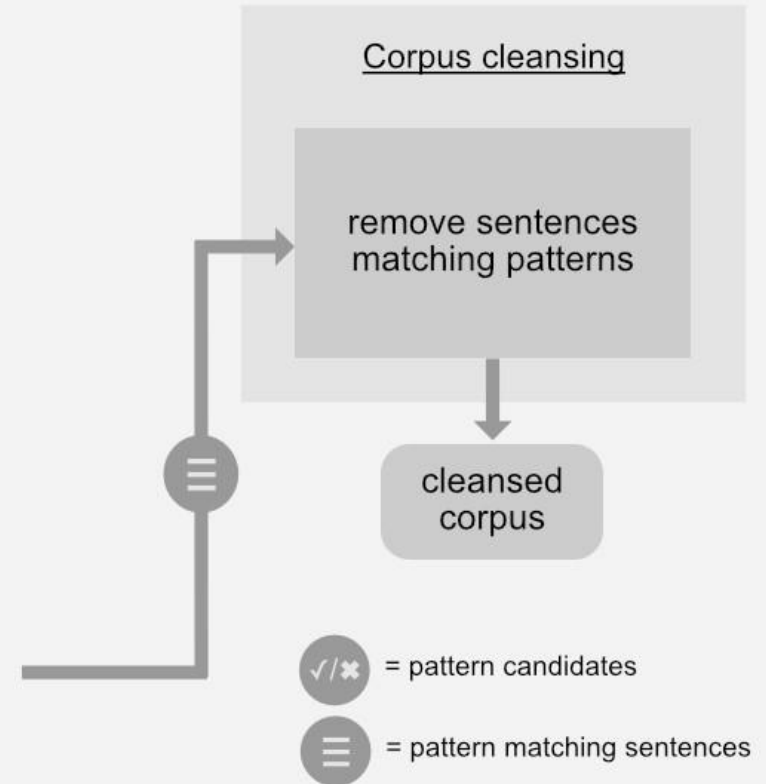
Approach – Concept



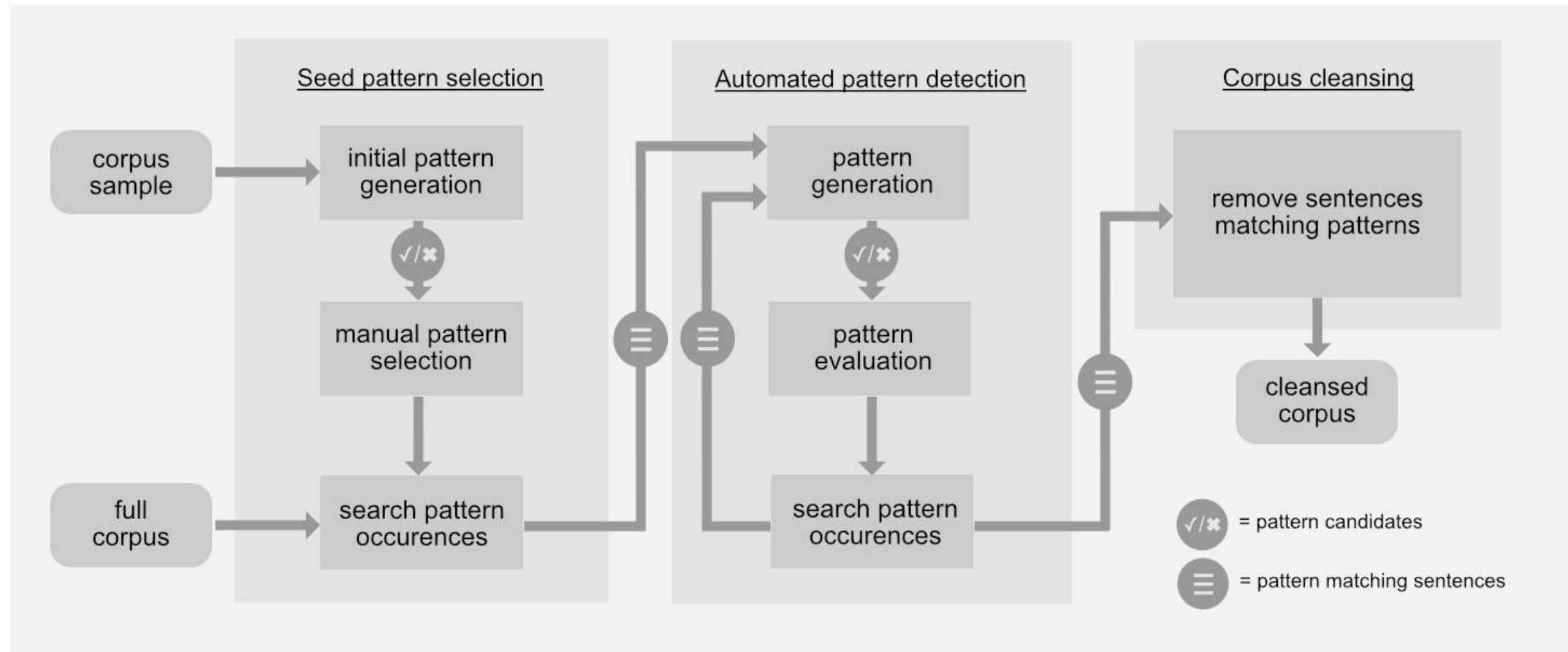
Approach – Concept



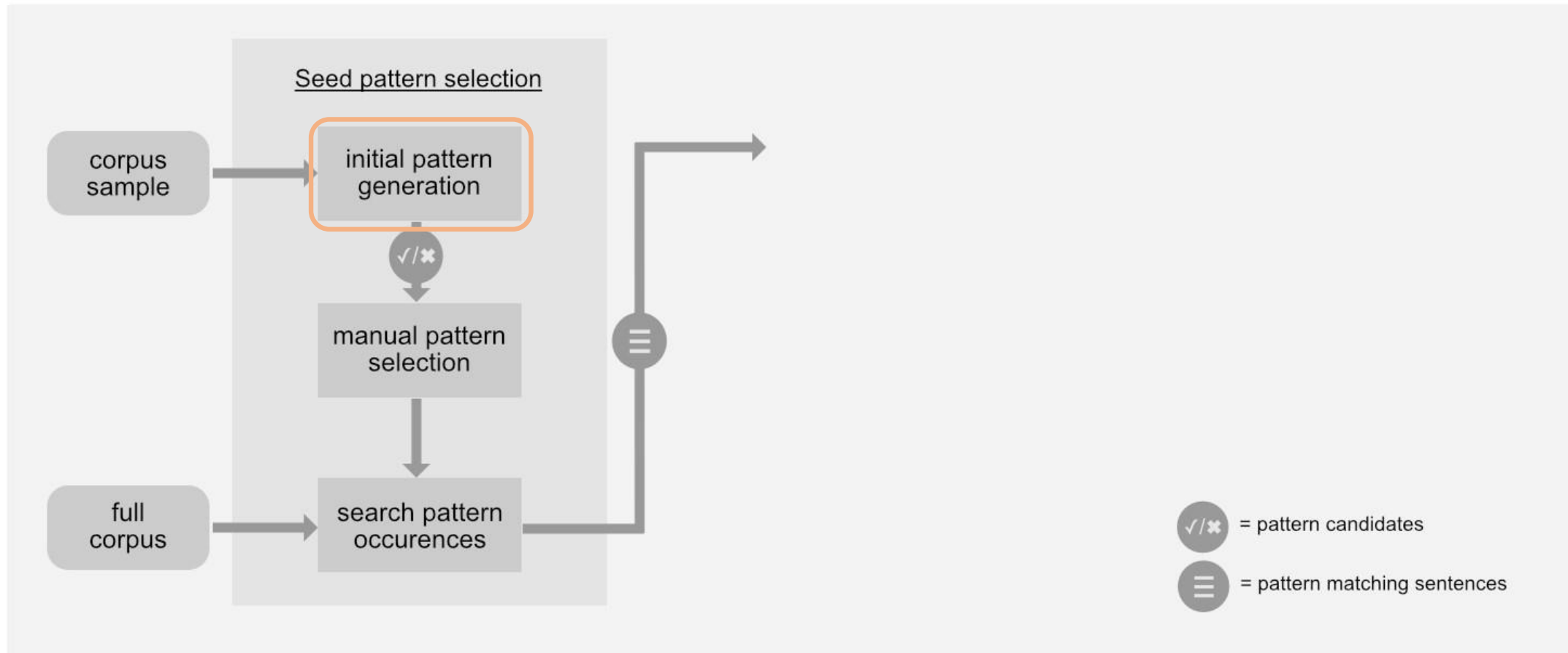
Approach – Concept



Approach – Concept



Approach – Seed pattern selection



Approach – Seed pattern selection

Pattern type	n-gram	Patterns of Relevant Sentences		Patterns of Irrelevant Sentences	
		Most frequent pattern	Score	Most frequent pattern	Score
Counts w/ Stopwords	1-gram	people	36 287	opponent	29 088
	2-gram	the world	6 210	my opponent	27 149
	3-gram	the fact that	4 206	my opponent s	3 983
	4-gram	in the united states	977	thank my opponent for	1251
	5-gram	has nothing to do with	494	i thank my opponent for	682
Counts w/o Stopwords	1-gram	people	36 287	opponent	29 088
	2-gram	united states	3 906	thank opponent	1 494
	3-gram	big bang theory	251	first round acceptance	617
	4-gram	life liberty pursuit happiness	102	would like thank opponent	359
	5-gram	make law respecting establishment religion	69	round 1 acceptance round 2	82
TF-IDF w/ Stopwords	1-gram	chronicled	1.00	—	—
	2-gram	and weaponry	0.56	actually forfeiting	1.00
	3-gram	an infinite regression	1.00	—	—
	4-gram	abusive education and domestic	1.00	—	—
	5-gram	acceptance of metapyhsical space that	1.00	—	—
TF-IDF w/o Stopwords	1-gram	abortions	1.00	—	—
	2-gram	americans like	1.00	—	—
	3-gram	able kill others	1.00	—	—
	4-gram	accidentally killing equivalent purposely	1.00	—	—
	5-gram	able disprove evolution instead creationists	1.00	—	—

Approach – Seed pattern selection

Pattern type	n-gram	Patterns of Relevant Sentences		Patterns of Irrelevant Sentences	
		Most frequent pattern	Score	Most frequent pattern	Score
Counts w/ Stopwords	1-gram	people	36 287	opponent	29 088
	2-gram	the world	6 210	my opponent	27 149
	3-gram	the fact that	4 206	my opponent s	3 983
	4-gram	in the united states	977	thank my opponent for	1251
	5-gram	has nothing to do with	494	i thank my opponent for	682
Counts w/o Stopwords	1-gram	people	36 287	opponent	29 088
	2-gram	united states	3 906	thank opponent	1 494
	3-gram	big bang theory	251	first round acceptance	617
	4-gram	life liberty pursuit happiness	102	would like thank opponent	359
	5-gram	make law respecting establishment religion	69	round 1 acceptance round 2	82
TF-IDF w/ Stopwords	1-gram	chronicled	1.00	—	—
	2-gram	and weaponry	0.56	actually forfeiting	1.00
	3-gram	an infinite regression	1.00	—	—
	4-gram	abusive education and domestic	1.00	—	—
	5-gram	acceptance of metapyhsical space that	1.00	—	—
TF-IDF w/o Stopwords	1-gram	abortions	1.00	—	—
	2-gram	americans like	1.00	—	—
	3-gram	able kill others	1.00	—	—
	4-gram	accidentally killing equivalent purposely	1.00	—	—
	5-gram	able disprove evolution instead creationists	1.00	—	—

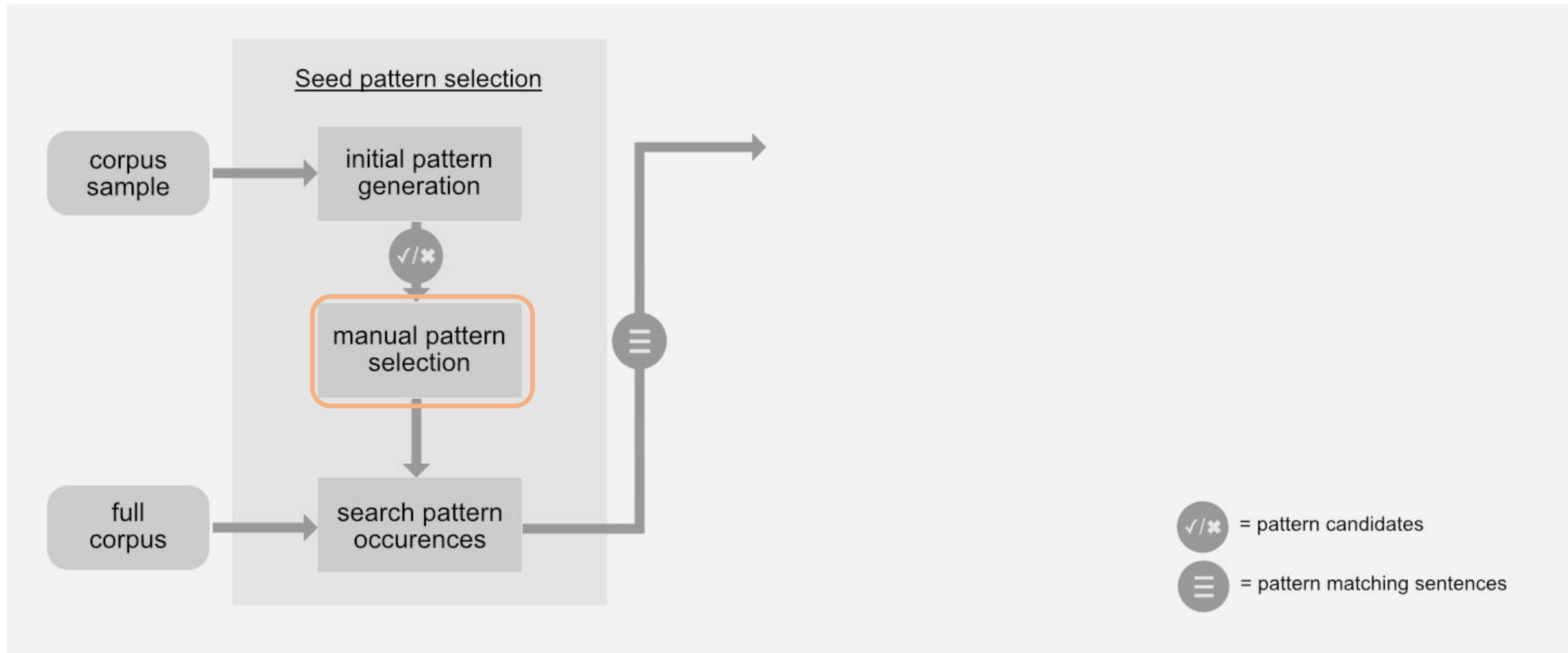
Approach – Seed pattern selection

Pattern type	n-gram	Patterns of Relevant Sentences		Patterns of Irrelevant Sentences	
		Most frequent pattern	Score	Most frequent pattern	Score
Counts w/ Stopwords	1-gram	people	36 287	opponent	29 088
	2-gram	the world	6 210	my opponent	27 149
	3-gram	the fact that	4 206	my opponent s	3 983
	4-gram	in the united states	977	thank my opponent for	1251
	5-gram	has nothing to do with	494	i thank my opponent for	682
Counts w/o Stopwords	1-gram	people	36 287	opponent	29 088
	2-gram	united states	3 906	thank opponent	1 494
	3-gram	big bang theory	251	first round acceptance	617
	4-gram	life liberty pursuit happiness	102	would like thank opponent	359
	5-gram	make law respecting establishment religion	69	round 1 acceptance round 2	82
TF-IDF w/ Stopwords	1-gram	chronicled	1.00	—	—
	2-gram	and weaponry	0.56	actually forfeiting	1.00
	3-gram	an infinite regression	1.00	—	—
	4-gram	abusive education and domestic	1.00	—	—
	5-gram	acceptance of metapyhsical space that	1.00	—	—
TF-IDF w/o Stopwords	1-gram	abortions	1.00	—	—
	2-gram	americans like	1.00	—	—
	3-gram	able kill others	1.00	—	—
	4-gram	accidentally killing equivalent purposely	1.00	—	—
	5-gram	able disprove evolution instead creationists	1.00	—	—

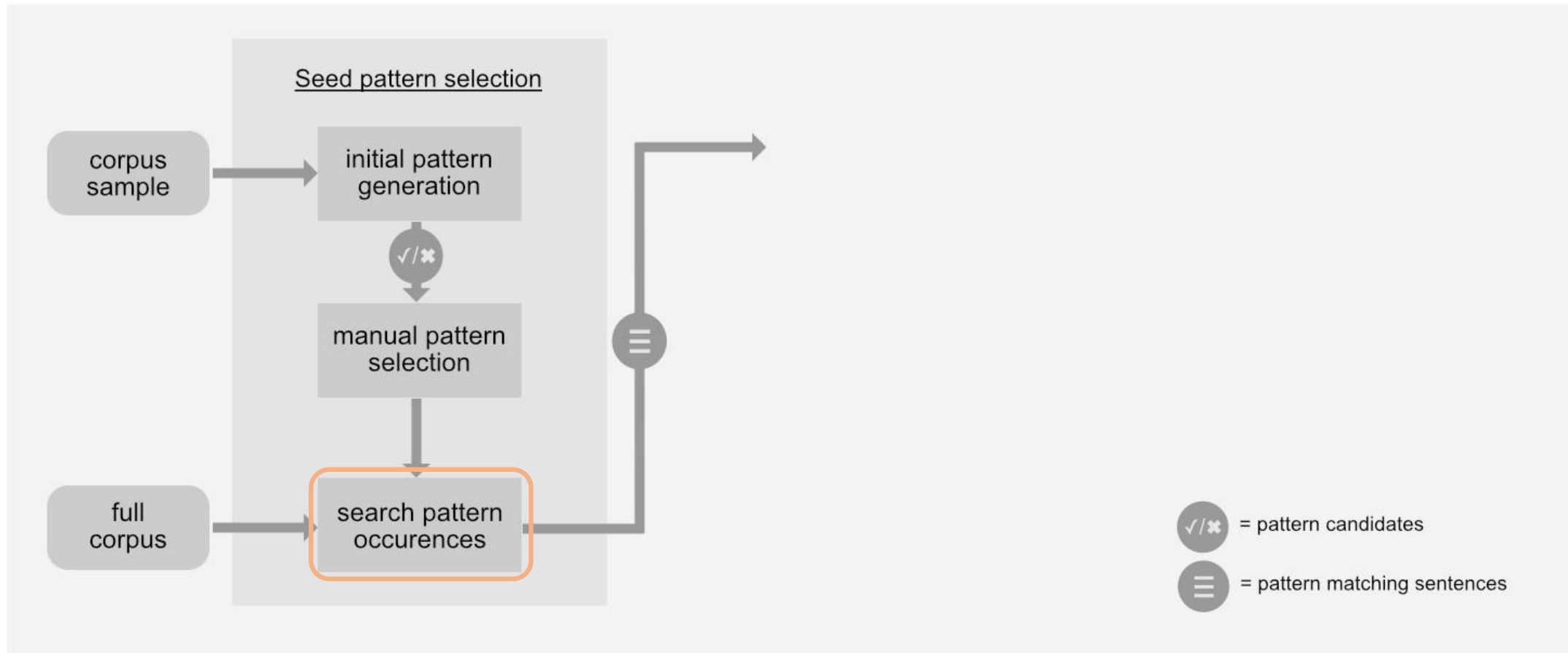
Approach – Seed pattern selection

Pattern type	n-gram	Patterns of Relevant Sentences		Patterns of Irrelevant Sentences	
		Most frequent pattern	Score	Most frequent pattern	Score
Counts w/ Stopwords	1-gram	people	36 287	opponent	29 088
	2-gram	the world	6 210	my opponent	27 149
	3-gram	the fact that	4 206	my opponent s	3 983
	4-gram	in the united states	977	thank my opponent for	1251
	5-gram	has nothing to do with	494	i thank my opponent for	682
Counts w/o Stopwords	1-gram	people	36 287	opponent	29 088
	2-gram	united states	3 906	thank opponent	1 494
	3-gram	big bang theory	251	first round acceptance	617
	4-gram	life liberty pursuit happiness	102	would like thank opponent	359
	5-gram	make law respecting establishment religion	69	round 1 acceptance round 2	82
TF-IDF w/ Stopwords	1-gram	chronicled	1.00	—	—
	2-gram	and weaponry	0.56	actually forfeiting	1.00
	3-gram	an infinite regression	1.00	—	—
	4-gram	abusive education and domestic	1.00	—	—
	5-gram	acceptance of metapyhsical space that	1.00	—	—
TF-IDF w/o Stopwords	1-gram	abortions	1.00	—	—
	2-gram	americans like	1.00	—	—
	3-gram	able kill others	1.00	—	—
	4-gram	accidentally killing equivalent purposely	1.00	—	—
	5-gram	able disprove evolution instead creationists	1.00	—	—

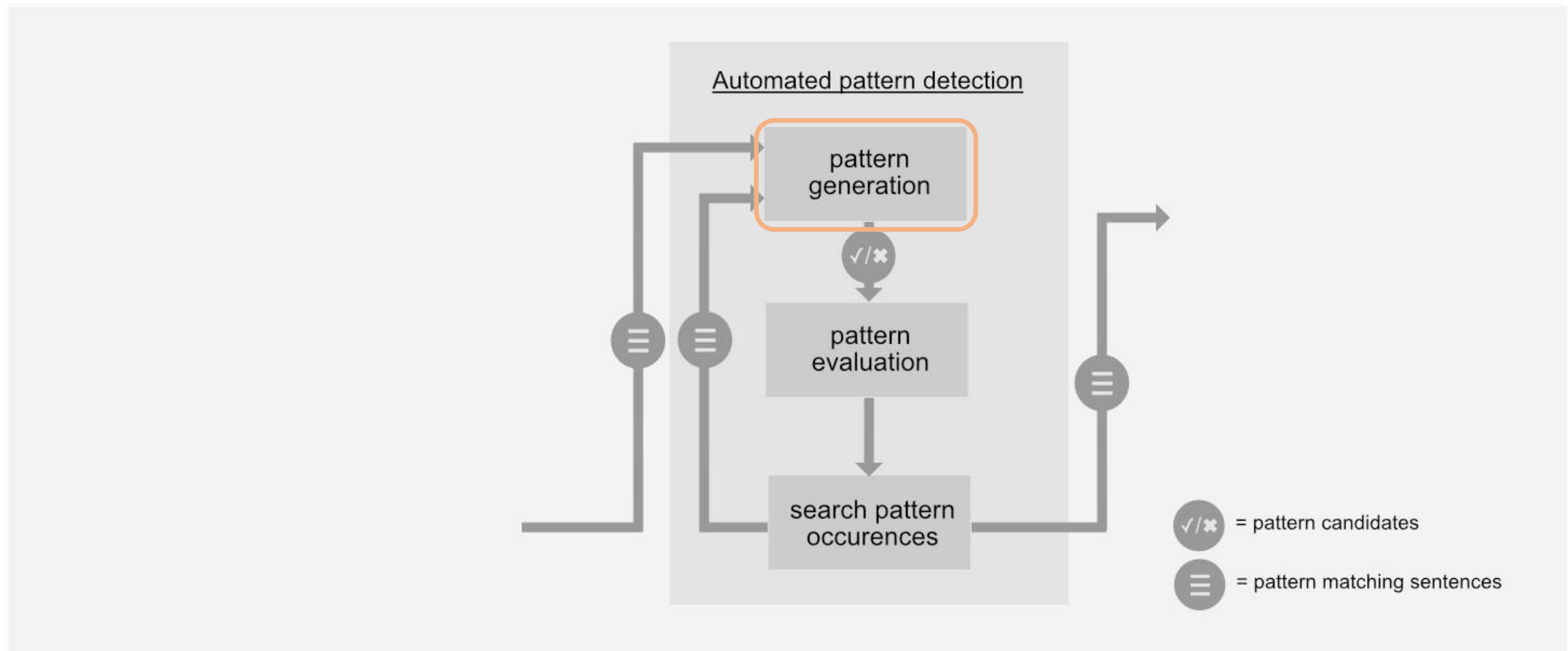
Approach – Seed pattern selection



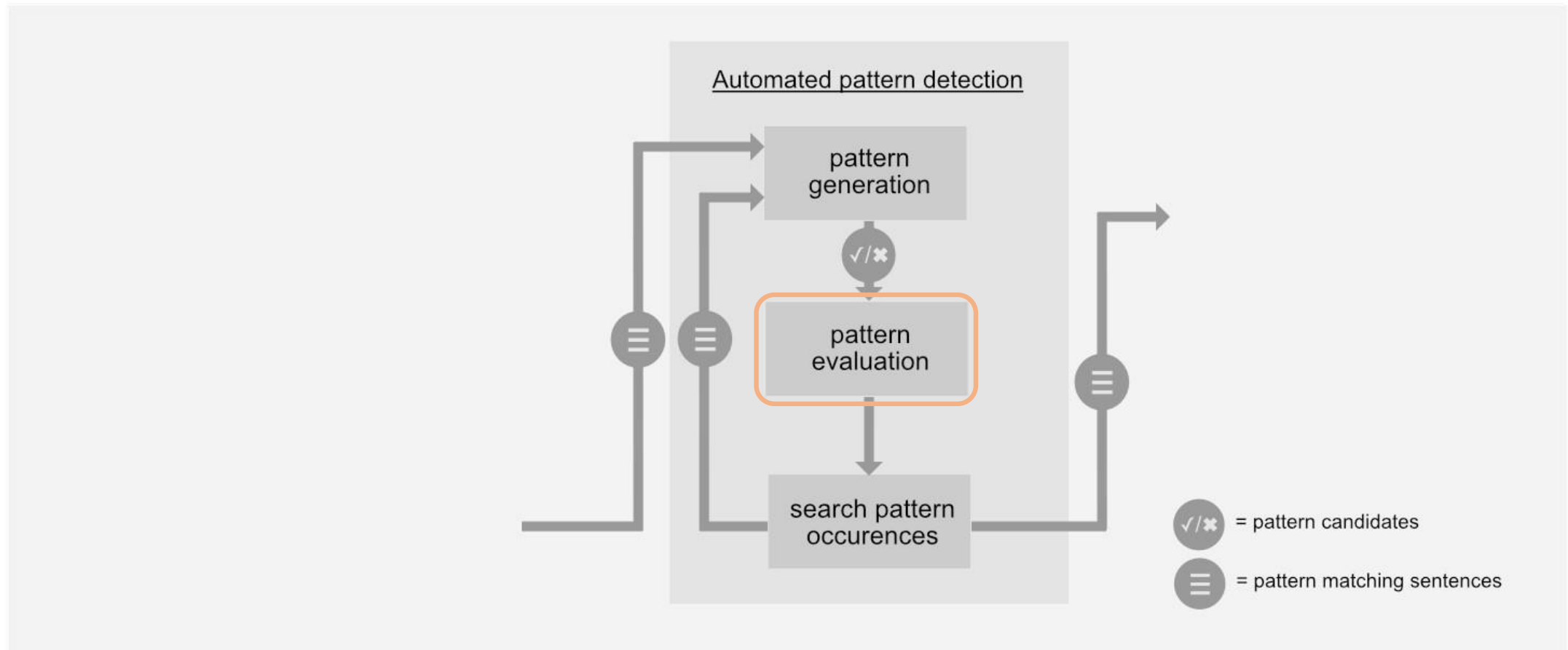
Approach – Seed pattern selection



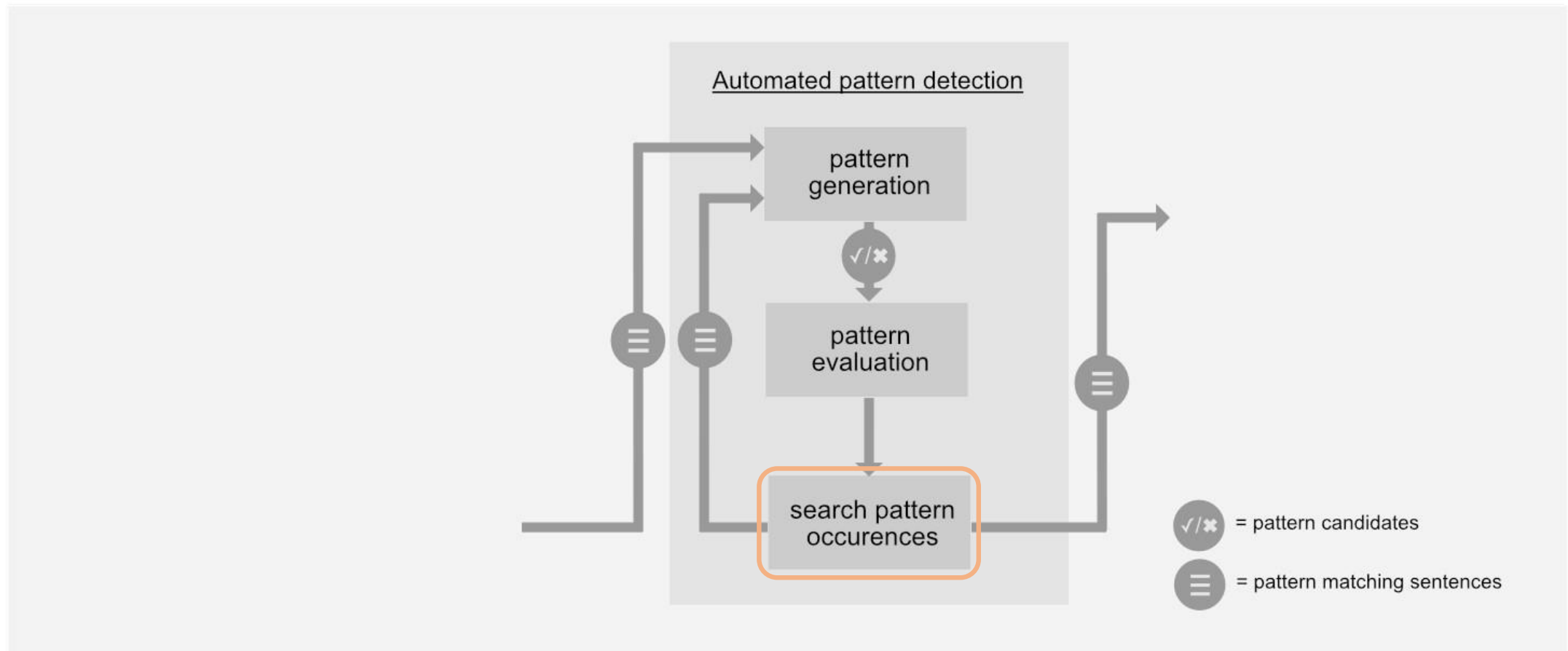
Approach – Automated pattern detection



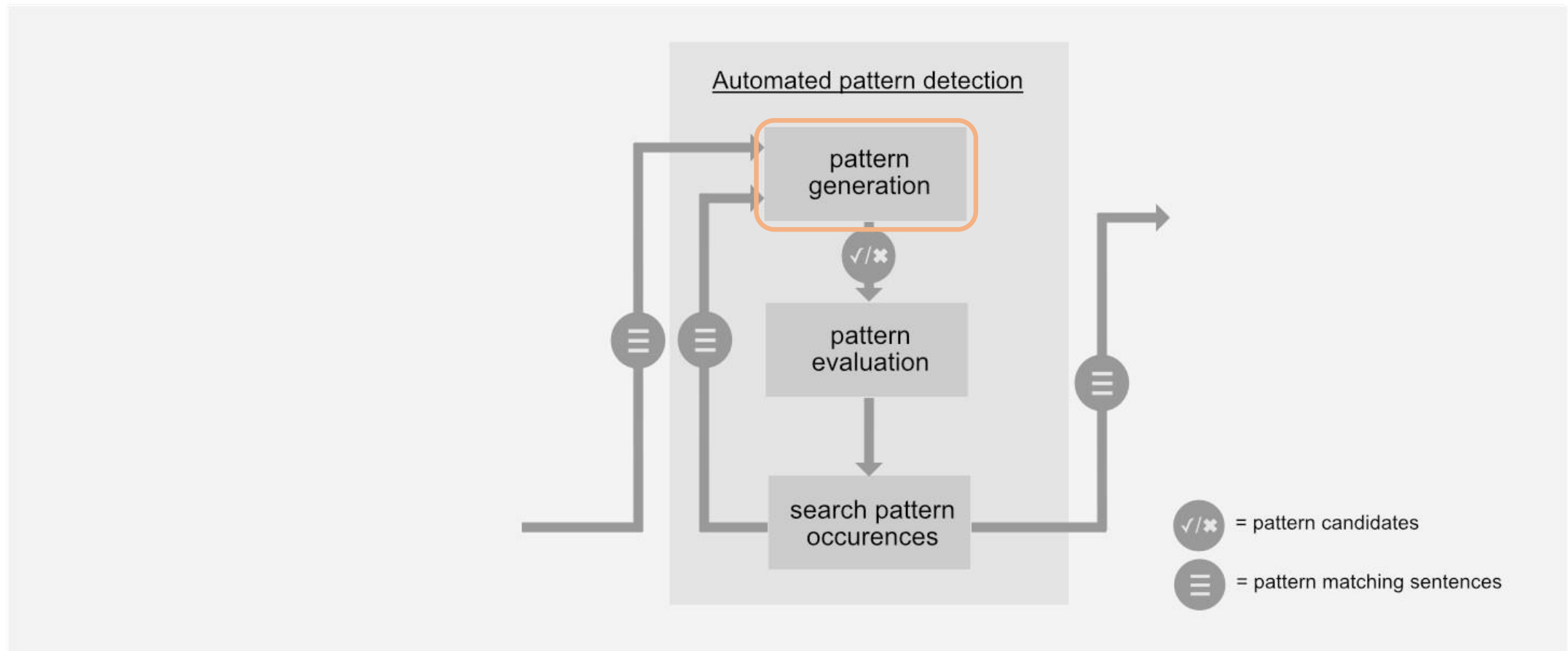
Approach – Automated pattern detection



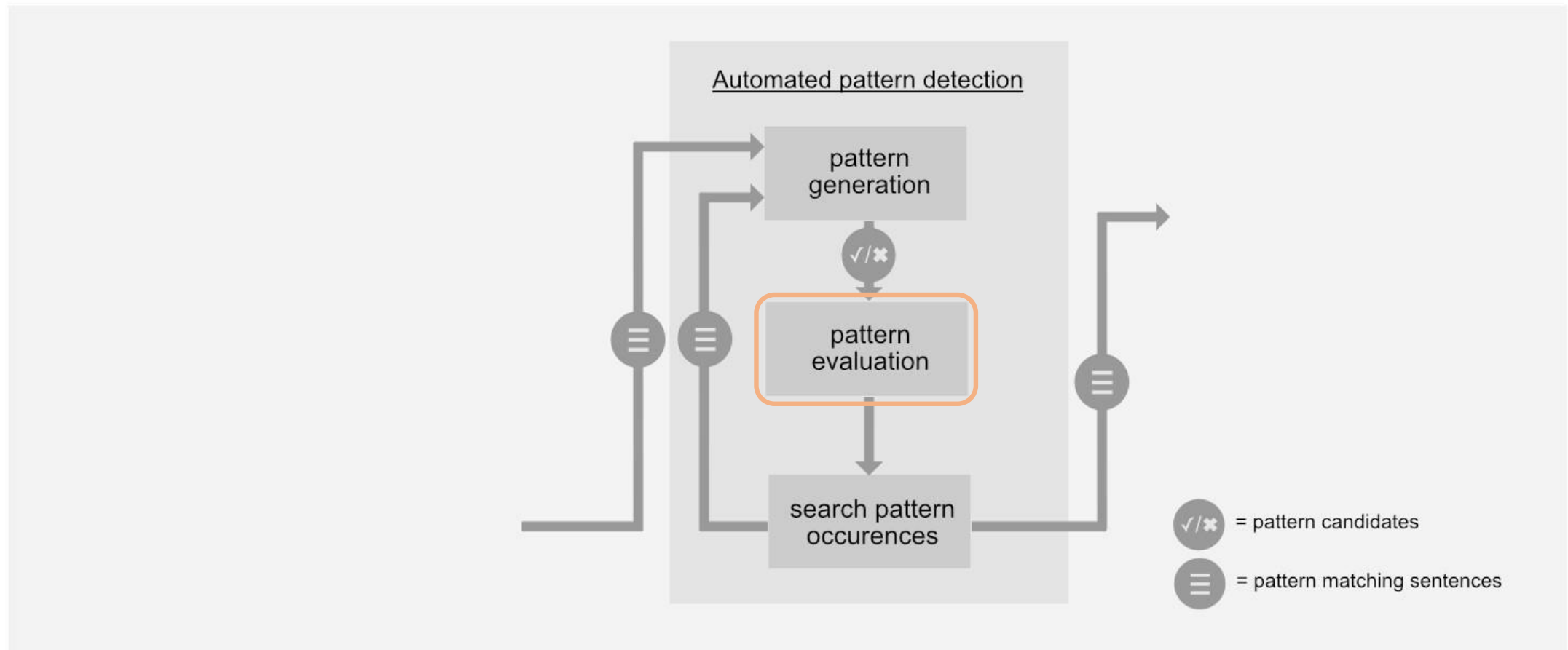
Approach – Automated pattern detection



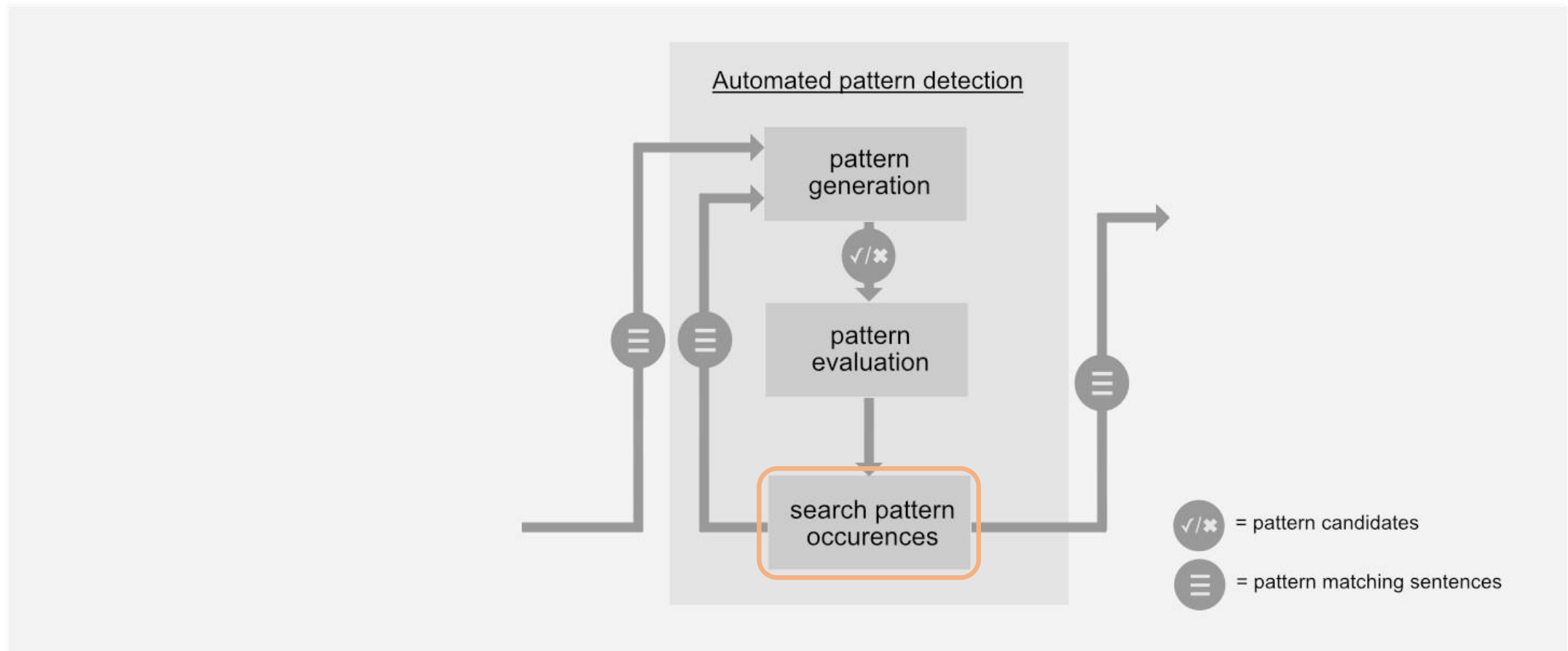
Approach – Automated pattern detection



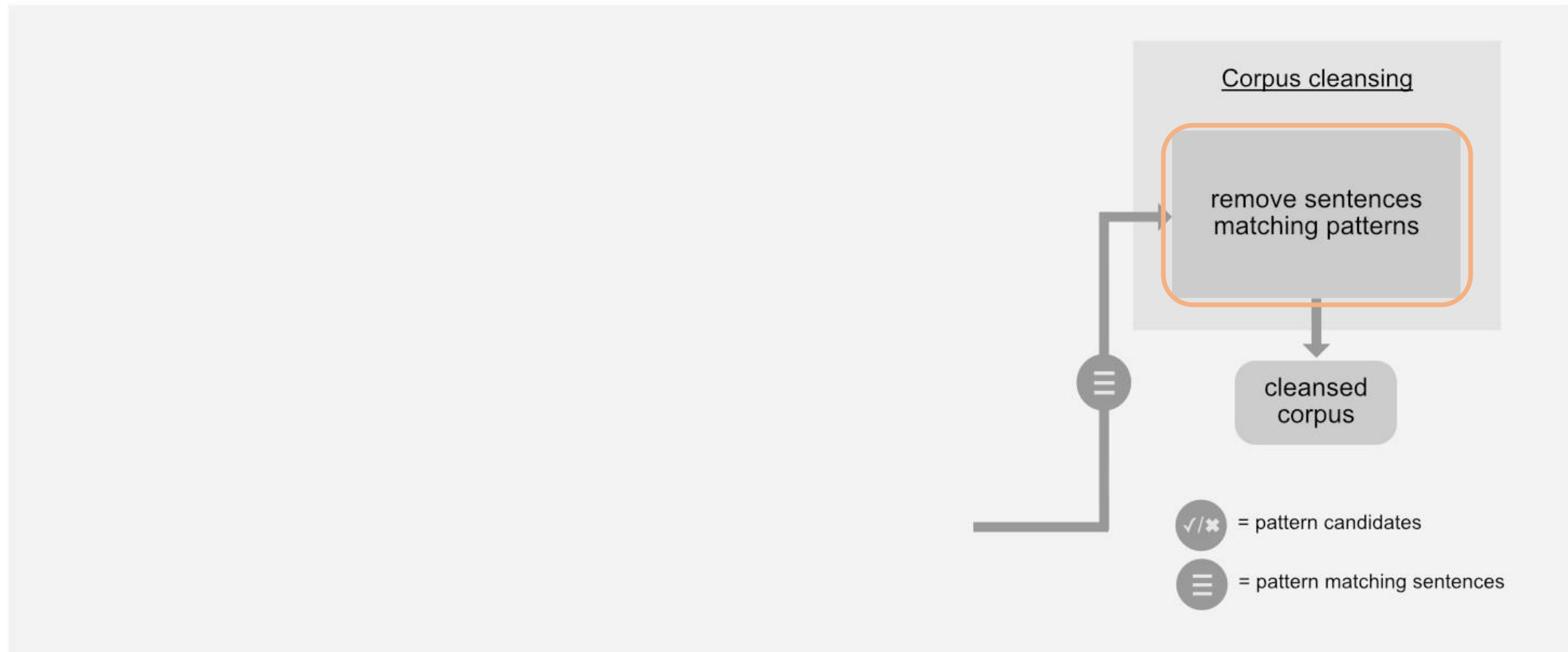
Approach – Automated pattern detection



Approach – Automated pattern detection



Approach – Corpus cleansing



Approach – Corpus cleansing

Apply special metrics?

Approach – Corpus cleansing

Apply special metrics?

Only remove sentences clearly irrelevant.

Approach – Corpus cleansing

Apply special metrics?

Only remove sentences clearly irrelevant.

Assumption: Most irrelevant sentences are located around the beginning and end of an argument.

Evaluation – Automated process

Iteration	Positive Patterns			
	Patterns	Sentences	Auto Prec.	Manual Prec.
Seed	38	600469	0.997	n/a
1	10	7602	0.987	n/a
2	0	-57	n/a	n/a
3	...	-15	...	n/a
4	...	-10	...	n/a
5	...	-6	...	n/a
Total	48	607983	0.984	n/a

Iteration	Negative Patterns			
	Patterns	Sentences	Auto Prec.	Manual Prec.
Seed	17	41619	0.972	1.0
1	74	5849	0.983	1.0
2	19	3606	0.982	1.0
3	4	956	0.973	0.96
4	3	594	0.980	0.97
5	5	225	0.983	0.88
Total	122	52849	0.980	0.97

Evaluation – Automated process

Iteration	Positive Patterns			
	Patterns	Sentences	Auto Prec.	Manual Prec.
Seed	38	600469	0.997	n/a
1	10	7602	0.987	n/a
2	0	-57	n/a	n/a
3	...	-15	...	n/a
4	...	-10	...	n/a
5	...	-6	...	n/a
Total	48	607983	0.984	n/a

Iteration	Negative Patterns			
	Patterns	Sentences	Auto Prec.	Manual Prec.
Seed	17	41619	0.972	1.0
1	74	5849	0.983	1.0
2	19	3606	0.982	1.0
3	4	956	0.973	0.96
4	3	594	0.980	0.97
5	5	225	0.983	0.88
Total	122	52849	0.980	0.97

Evaluation – Automated process

Positive Patterns				
Iteration	Patterns	Sentences	Auto Prec.	Manual Prec.
Seed	38	600469	0.997	n/a
1	10	7602	0.987	n/a
2	0	-57	n/a	n/a
3	...	-15	...	n/a
4	...	-10	...	n/a
5	...	-6	...	n/a
Total	48	607983	0.984	n/a

Negative Patterns				
Iteration	Patterns	Sentences	Auto Prec.	Manual Prec.
Seed	17	41619	0.972	1.0
1	74	5849	0.983	1.0
2	19	3606	0.982	1.0
3	4	956	0.973	0.96
4	3	594	0.980	0.97
5	5	225	0.983	0.88
Total	122	52849	0.980	0.97

Evaluation – Automated process

Positive Patterns				
Iteration	Patterns	Sentences	Auto Prec.	Manual Prec.
Seed	38	600469	0.997	n/a
1	10	7602	0.987	n/a
2	0	-57	n/a	n/a
3	...	-15	...	n/a
4	...	-10	...	n/a
5	...	-6	...	n/a
Total	48	607983	0.984	n/a

Negative Patterns				
Iteration	Patterns	Sentences	Auto Prec.	Manual Prec.
Seed	17	41619	0.972	1.0
1	74	5849	0.983	1.0
2	19	3606	0.982	1.0
3	4	956	0.973	0.96
4	3	594	0.980	0.97
5	5	225	0.983	0.88
Total	122	52849	0.980	0.97

Evaluation – Automated process

Positive Patterns				
Iteration	Patterns	Sentences	Auto Prec.	Manual Prec.
Seed	38	600469	0.997	n/a
1	10	7602	0.987	n/a
2	0	-57	n/a	n/a
3	...	-15	...	n/a
4	...	-10	...	n/a
5	...	-6	...	n/a
Total	48	607983	0.984	n/a

Negative Patterns				
Iteration	Patterns	Sentences	Auto Prec.	Manual Prec.
Seed	17	41619	0.972	1.0
1	74	5849	0.983	1.0
2	19	3606	0.982	1.0
3	4	956	0.973	0.96
4	3	594	0.980	0.97
5	5	225	0.983	0.88
Total	122	52849	0.980	0.97

Evaluation – Manual Annotation

- Three experts
- 100 Sentences per round
+100 Sentences from seed

Iteration	Positive Patterns			
	Patterns	Sentences	Auto Prec.	Manual Prec.
Seed	38	600469	0.997	n/a
1	10	7602	0.987	n/a
2	0	-57	n/a	n/a
3	...	-15	...	n/a
4	...	-10	...	n/a
5	...	-6	...	n/a
Total	48	607983	0.984	n/a

Iteration	Negative Patterns			
	Patterns	Sentences	Auto Prec.	Manual Prec.
Seed	17	41619	0.972	1.0
1	74	5849	0.983	1.0
2	19	3606	0.982	1.0
3	4	956	0.973	0.96
4	3	594	0.980	0.97
5	5	225	0.983	0.88
Total	122	52849	0.980	0.97

Evaluation – Manual Annotation

Iteration	Annotator Precision					
	Individual			Combined		
	1	2	3	Full	Majority	At least one
Seed	1.0	0.99	1.0	0.99	1.0	1.0
1	0.98	0.98	1.0	0.96	1.0	1.0
2	1.0	0.95	0.99	0.94	1.0	1.0
3	0.96	0.96	0.95	0.93	0.96	0.98
4	0.95	0.95	0.97	0.93	0.97	0.97
5	0.84	0.85	0.93	0.79	0.88	0.95
Total	0.955	0.946	0.973	0.92	0.97	0.98

Evaluation – Manual Annotation

Iteration	Annotator Precision					
	Individual			Combined		
	1	2	3	Full	Majority	At least one
Seed	1.0	0.99	1.0	0.99	1.0	1.0
1	0.98	0.98	1.0	0.96	1.0	1.0
2	1.0	0.95	0.99	0.94	1.0	1.0
3	0.96	0.96	0.95	0.93	0.96	0.98
4	0.95	0.95	0.97	0.93	0.97	0.97
5	0.84	0.85	0.93	0.79	0.88	0.95
Total	0.955	0.946	0.973	0.92	0.97	0.98

Evaluation – Manual Annotation

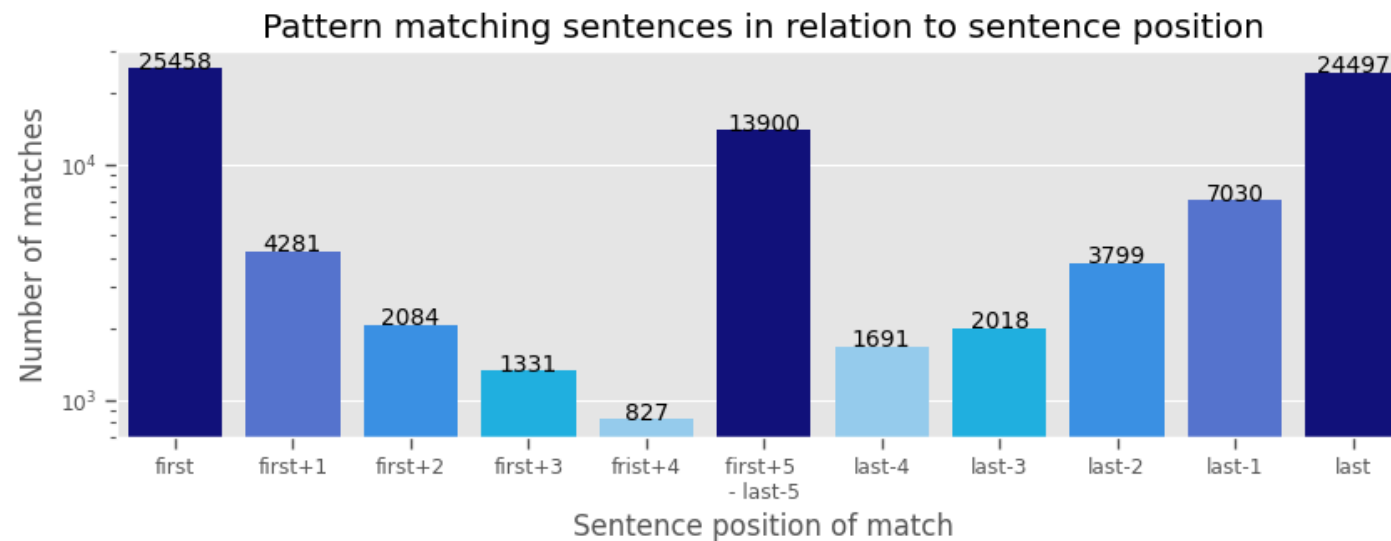
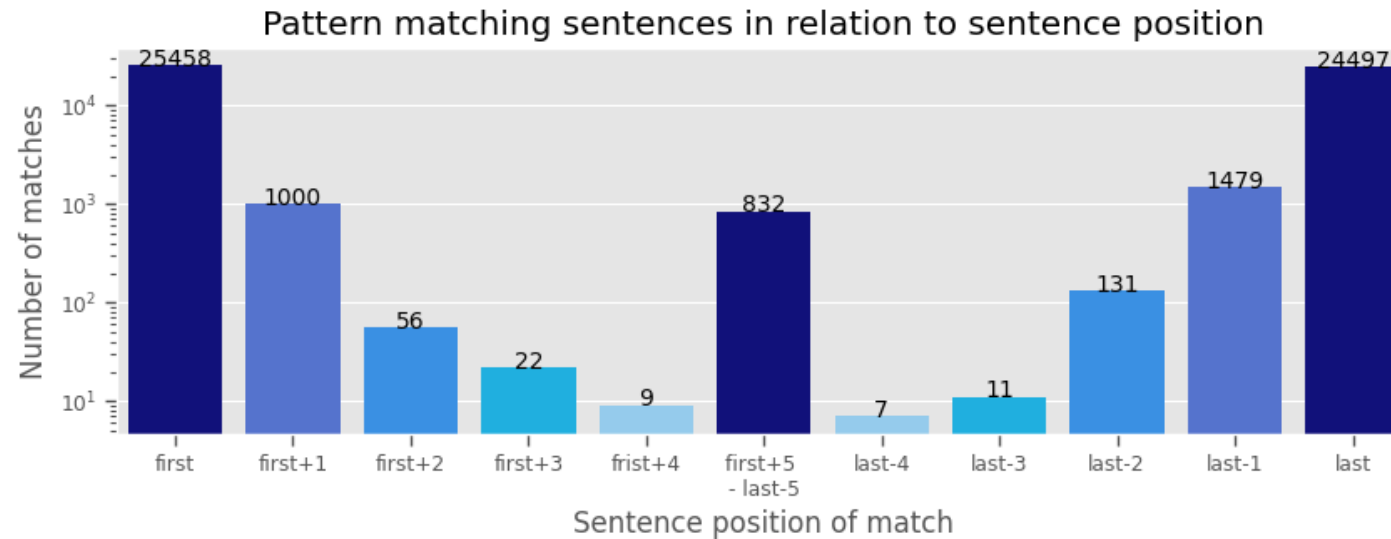
Iteration	Annotator Precision					
	Individual			Combined		
	1	2	3	Full	Majority	At least one
Seed	1.0	0.99	1.0	0.99	1.0	1.0
1	0.98	0.98	1.0	0.96	1.0	1.0
2	1.0	0.95	0.99	0.94	1.0	1.0
3	0.96	0.96	0.95	0.93	0.96	0.98
4	0.95	0.95	0.97	0.93	0.97	0.97
5	0.84	0.85	0.93	0.79	0.88	0.95
Total	0.955	0.946	0.973	0.92	0.97	0.98

Just remove every matched sentence?

Evaluation – Sentence Position

Assumption: Most irrelevant sentences are located around the beginning and end of an argument

Evaluation – Sentence Position



Conclusion – Overall performance

Overall Performance:

- 53 502 sentences from 48 089 arguments removed
-> more than 10% of arguments filtered (from complete corpus)

Conclusion – Overall performance

Overall Performance:

- 53 502 sentences from 48 089 arguments removed
 - > more then 10% of arguments filtered (from complete corpus)
- Average precision of 97% according to manual evaluation

Conclusion – Overall performance

Overall Performance:

- 53 502 sentences from 48 089 arguments removed
 - > more then 10% of arguments filtered (from complete corpus)
- Average precision of 97% according to manual evaluation

However: Some improvements and further analysis required

Conclusion – Future Improvements

Improvements:

- Improve tokenization

Conclusion – Future Improvements

Improvements:

- Improve tokenization
- Experiment with different pattern extraction methods

Conclusion – Future Improvements

Improvements:

- Improve tokenization
- Experiment with different pattern extraction methods
- Evaluate sentence position relevance

Thank you for your attention!

Any questions?

References

- Yamen Ajjour, Henning Wachsmuth, Johannes Kiesel, Martin Potthast, Matthias Hagen, and Benno Stein. Data acquisition for argument search: The args.me corpus. In Proceedings of the 42nd Edition of the German Conference on Artificial Intelligence, page 48â59, 2019. 2.3, 3, 4