Algorithms for the Manipulation of Writing Style

Martin Potthast

Leipzig University

webis.de

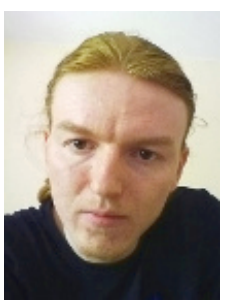
Outline

- Introduction
- Technology Basics
- Author Identification
- Constrained Paraphrasing
- Author Obfuscation

Joint work with

Benno Stein

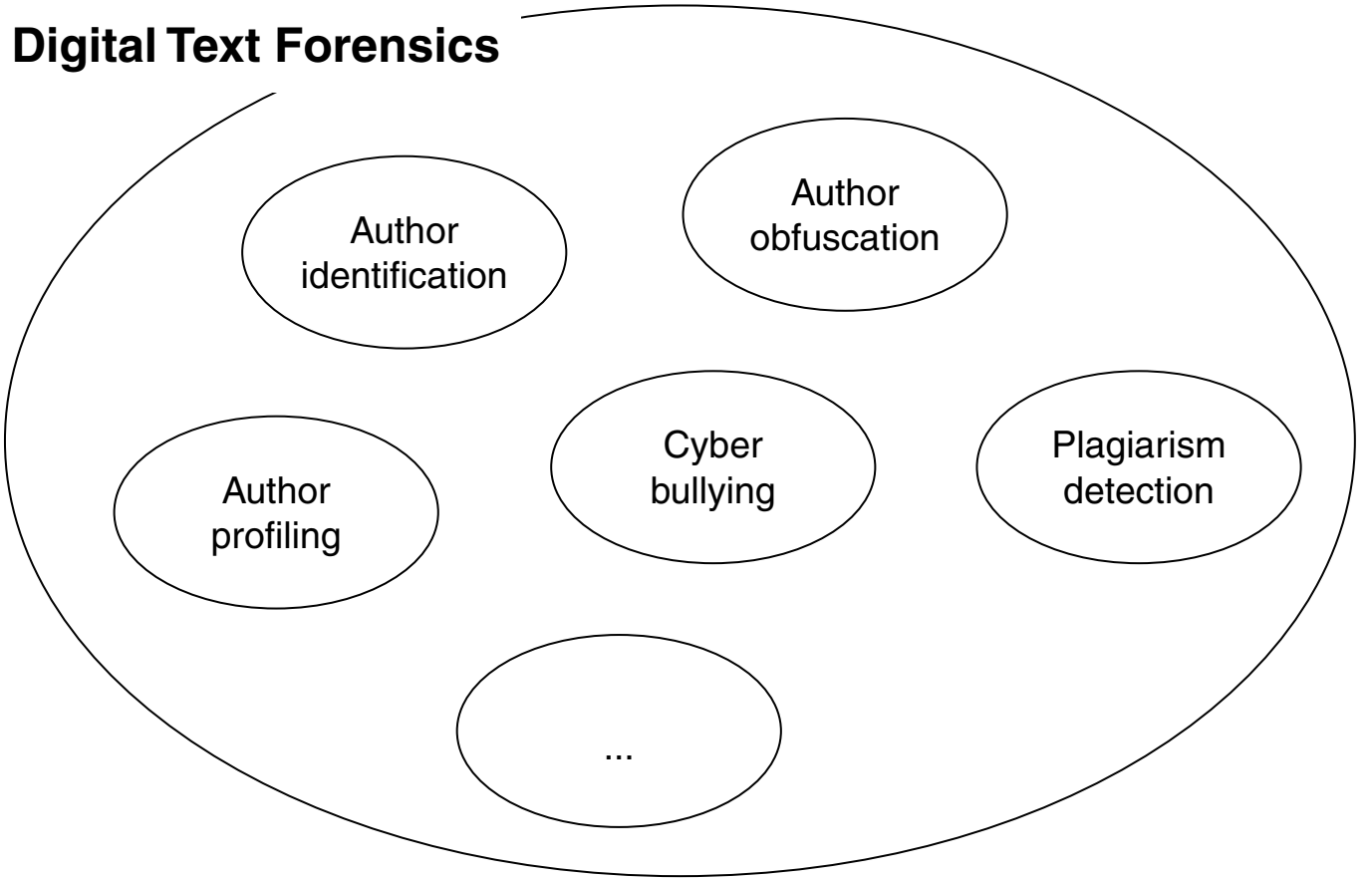
Matthias Hagen



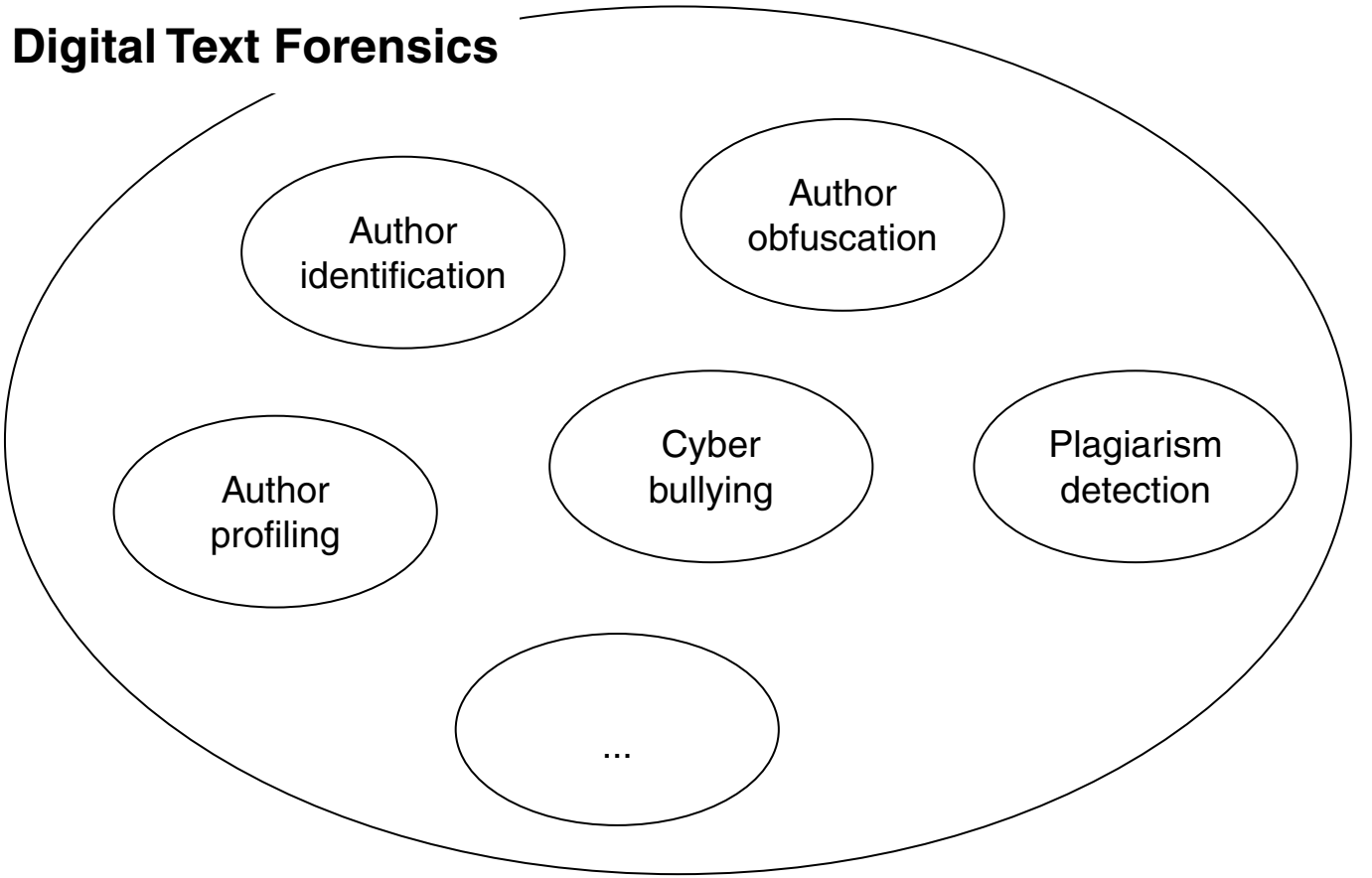
Janek Bevendorff

Christof Bräutigam

Digital Text Forensics



©Webis 2021



Retrieval models | Algorithms | Corpora

©Webis 2021

Some Technology Basics

Text with markup:

```
<TEXT> <TITLE>CHRYSLER> DEAL LEAVES UNCERTAINTY FOR AMC
WORKERS</TITLE> <AUTHOR> By Richard Walker, Reuters</AUTHOR>
<DATELINE> DETROIT, March 11 - </DATELINE><BODY>Chrysler
Corp's 1.5 billion dlr bid to takeover American Motors Corp;
AMO> should help bolster the small automaker's sales, but it
leaves the future of its 19,000 employees in doubt, industry
analysts say. It was "business as usual"yesterday at the
American ...
```


©Webis 2021

Raw text:

```
chrysler deal leaves uncertainty for amc workers by richard
walker reuters detroit march 11 chrysler corp s 1 5 billion
dlr bid to takeover american motors corp should help bolster
the small automaker s sales but it leaves the future of its
19 000 employees in doubt industry analysts say it was
business as usual yesterday at the american
```


©Webis 2021

Stop words:

chrysler deal leaves uncertainty **for** amc workers **by** richard walker reuters detroit **march 11** chrysler **corp s 1 5 billion dlr** bid **to** takeover american motors **corp should** help bolster **the small** automaker **s** sales **but it** leaves **the** future **of its 19 000** employees **in** doubt industry analysts **say it was** business **as usual** yesterday **at the** american

©Webis 2021

After stemming:

```
chrysler deal leav uncertain amc work richard walk reut
detroit takeover american motor help bols automak sal leav
futur employ doubt industr analy business usual yesterday
```


©Webis 2021

After stemming:

```
chrysler deal leav uncertain amc work richard walk reut
detroit takeover american motor help bols automak sal leav
futur employ doubt industr analy business usual yesterday
```

$$\mathbf{d} = \begin{pmatrix} \texttt{chrysler} & w_1 \\ \texttt{motor} & w_2 \\ \cdots & \\ \texttt{cat} & w_x \\ \texttt{dog} & w_y \\ \texttt{mouse} & w_z \end{pmatrix}$$

©Webis 2021

After stemming:

```
chrysler deal leav uncertain amc work richard walk reut
detroit takeover american motor help bols automak sal leav
futur employ doubt industr analy business usual yesterday
```

$$\mathbf{d} = \begin{pmatrix} \texttt{chrysler} & w_1 \\ \texttt{motor} & w_2 \\ \cdots & \\ \texttt{cat} & w_x \\ \texttt{dog} & w_y \\ \texttt{mouse} & w_z \end{pmatrix} \quad \rightsquigarrow \quad \begin{pmatrix} \texttt{chrysler} & 0.2 \\ \texttt{motor} & 0.3 \\ \cdots & \\ \texttt{cat} & 0.0 \\ \texttt{dog} & 0.1 \\ \texttt{mouse} & 0.1 \end{pmatrix}$$

Weight computation:

term frequency (tf), inverse document frequency (idf), divergence from randomness

©Webis 2021

After stemming:

```
chrysler deal leav uncertain amc work richard walk reut
detroit takeover american motor help bols automak sal leav
futur employ doubt industr analy business usual yesterday
```

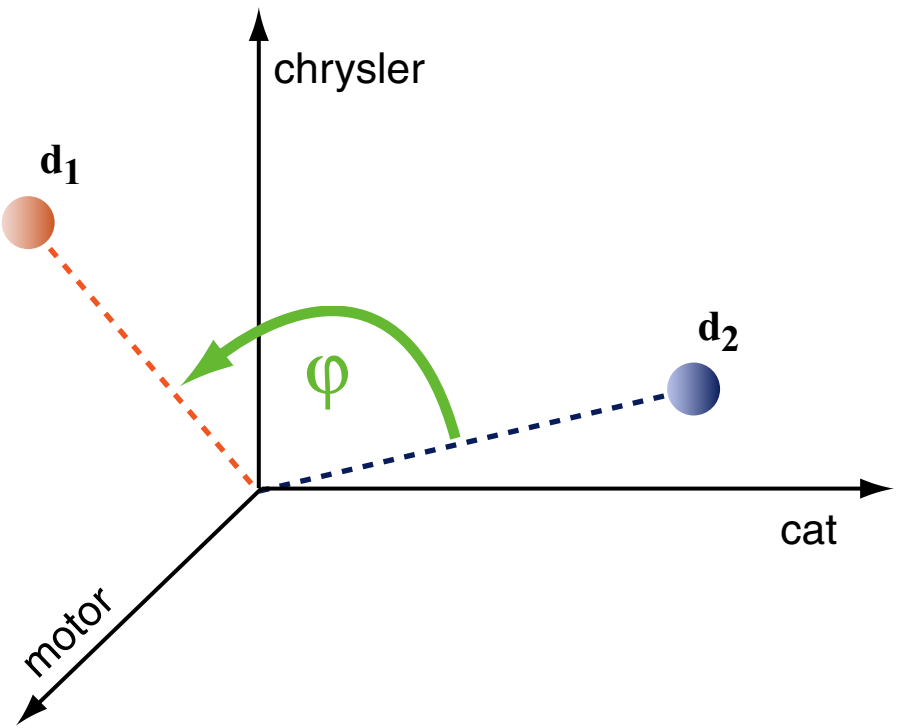
$$\mathbf{d} = \begin{pmatrix} \texttt{chrysler} & w_1 \\ \texttt{motor} & w_2 \\ \cdots & \\ \texttt{cat} & w_x \\ \texttt{dog} & w_y \\ \texttt{mouse} & w_z \end{pmatrix} \quad \rightsquigarrow \quad \left\langle \begin{pmatrix} \texttt{chrysler} & 0.2 \\ \texttt{motor} & 0.3 \\ \cdots & \\ \texttt{cat} & 0.0 \\ \texttt{dog} & 0.1 \\ \texttt{mouse} & 0.1 \end{pmatrix}, \begin{pmatrix} \texttt{chrysler} & 0.1 \\ \texttt{motor} & 0.2 \\ \cdots & \\ \texttt{cat} & 0.2 \\ \texttt{dog} & 0.0 \\ \texttt{mouse} & 0.0 \end{pmatrix} \right\rangle$$

Weight computation:

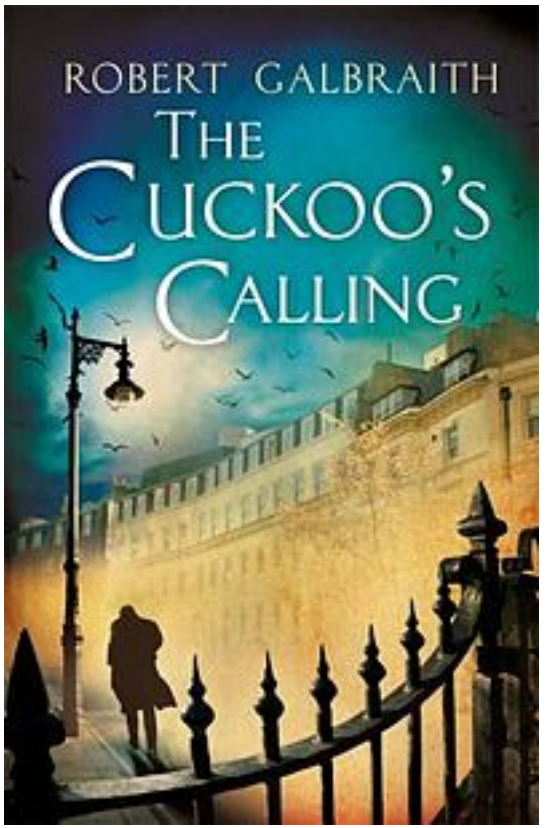
term frequency (tf), inverse document frequency (idf), divergence from randomness

©Webis 2021

Vector space:



©Webis 2021



Fake likes

Fake news

Fake clicks

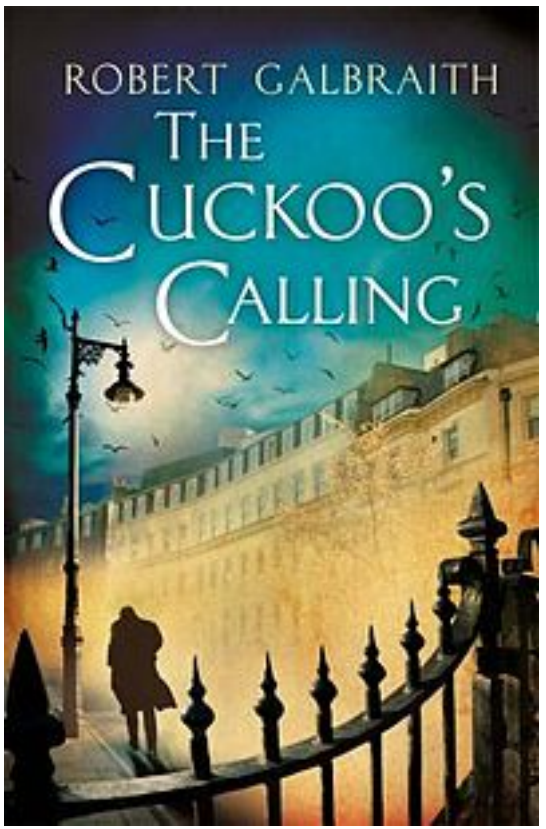
Fake users

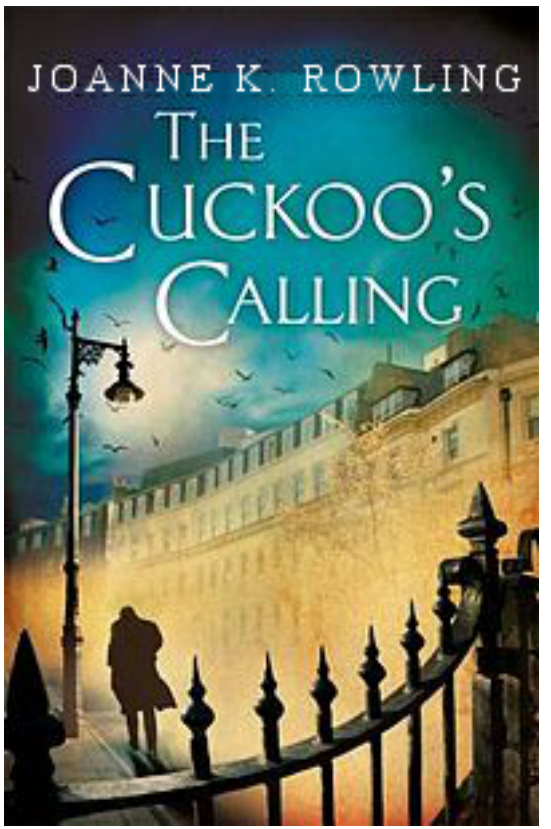
Fake reviews

Fake comments

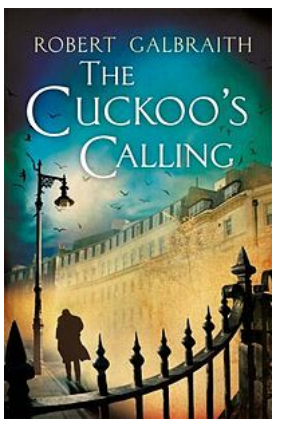
⋮

Fake identities (pseudonyms)

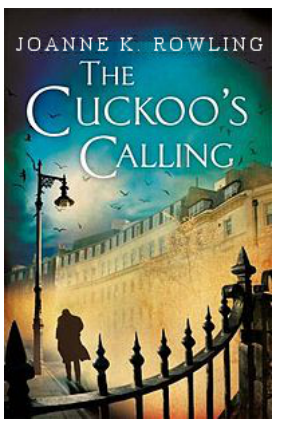

ROBERT GALBRAITH
THE CUCKOO'S CALLING


JOANNE K. ROWLING
THE CUCKOO'S CALLING

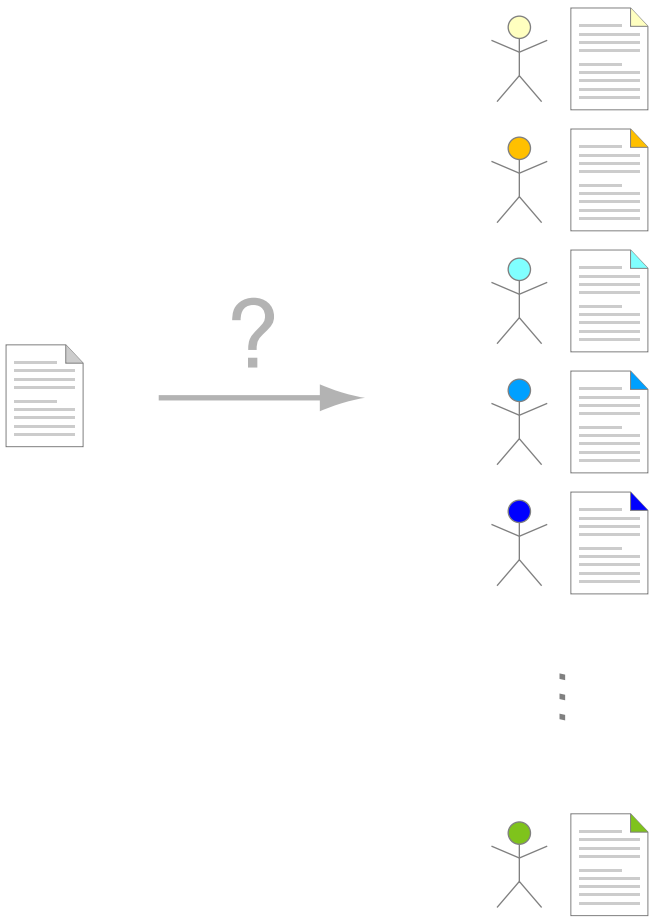
Author Identification



→



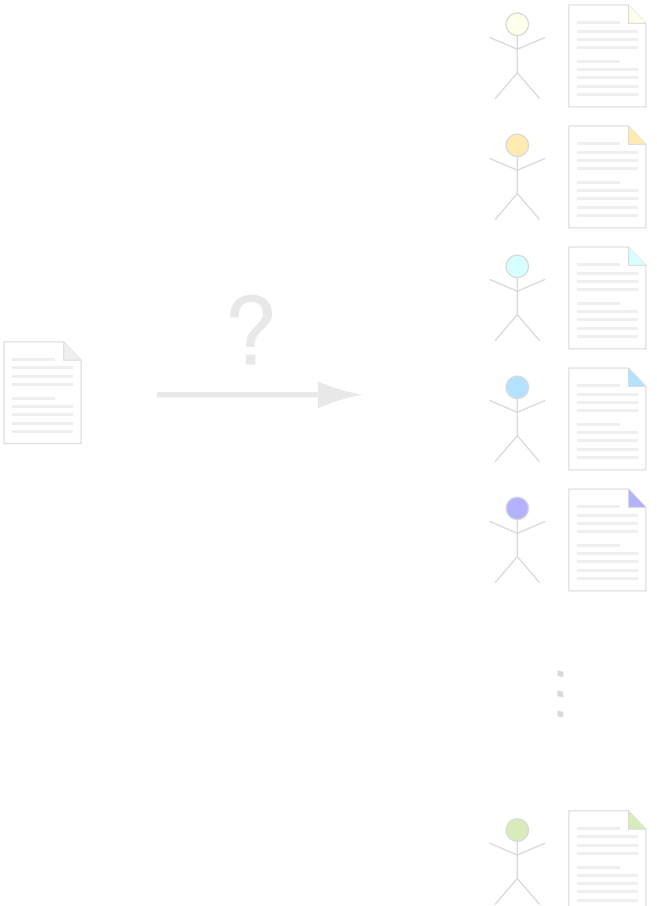
Authorship Attribution



To which author does a text belong?

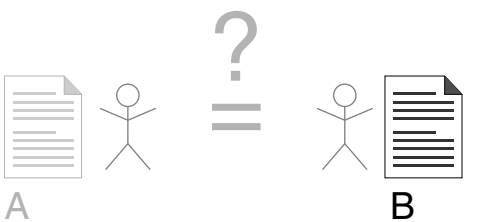
©Webis 2021

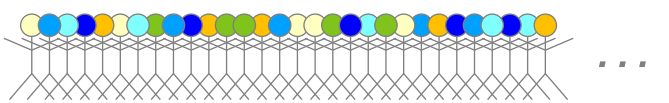
Authorship Attribution



To which author does a text belong?

Authorship Verification

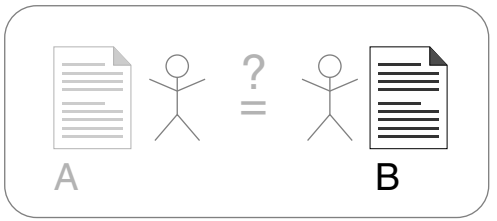




Originate two texts from the same author?

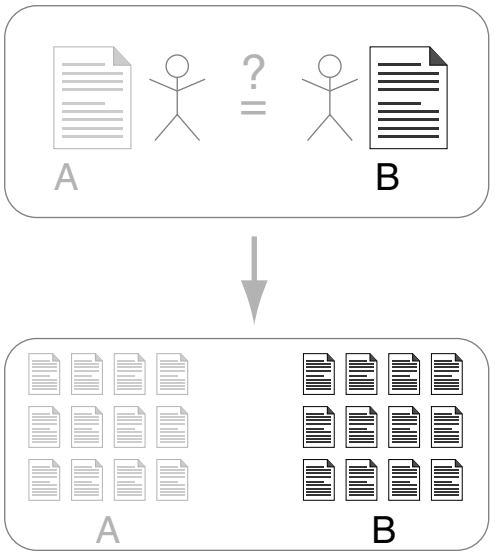
©Webis 2021

Authorship Verification via “Unmasking” [Koppel/Schler 2004]



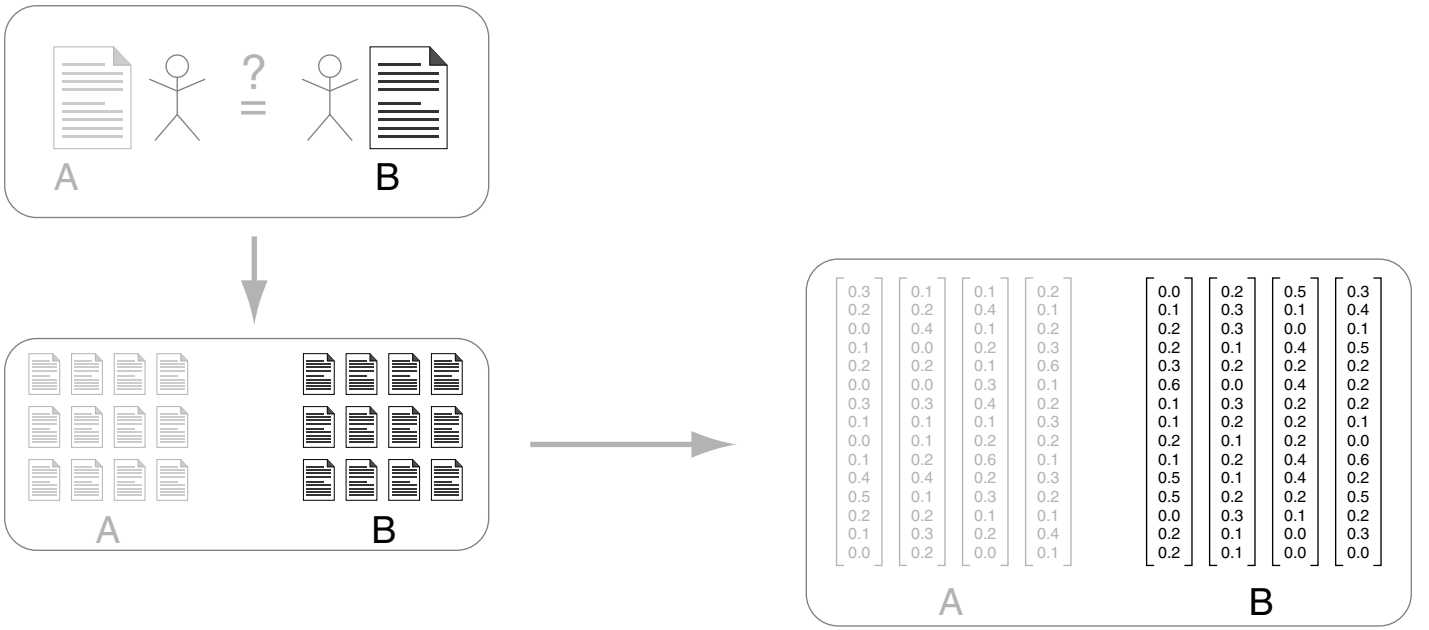
©Webis 2021

Authorship Verification via “Unmasking” [Koppel/Schler 2004]



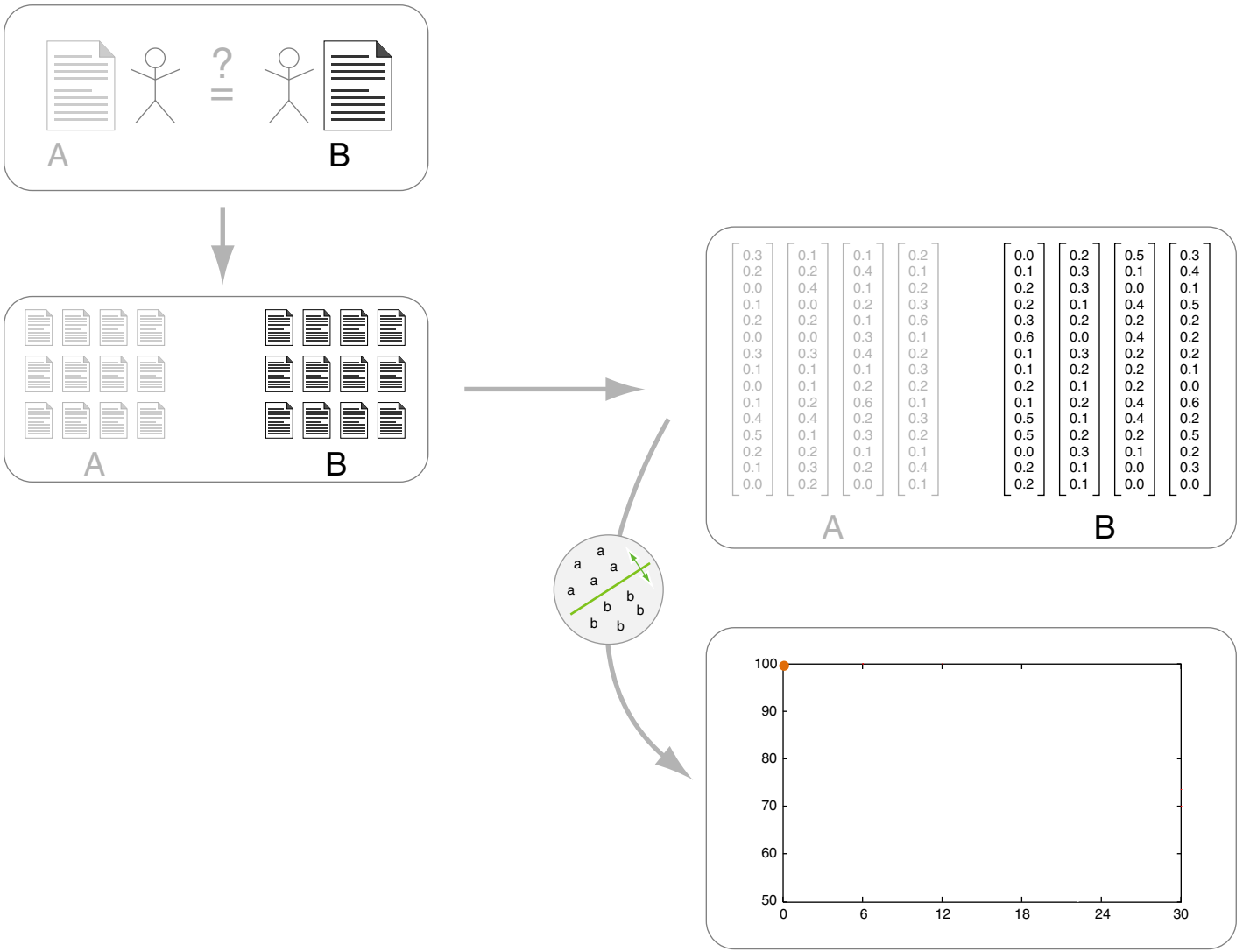
©Webis 2021

Authorship Verification via “Unmasking” [Koppel/Schler 2004]



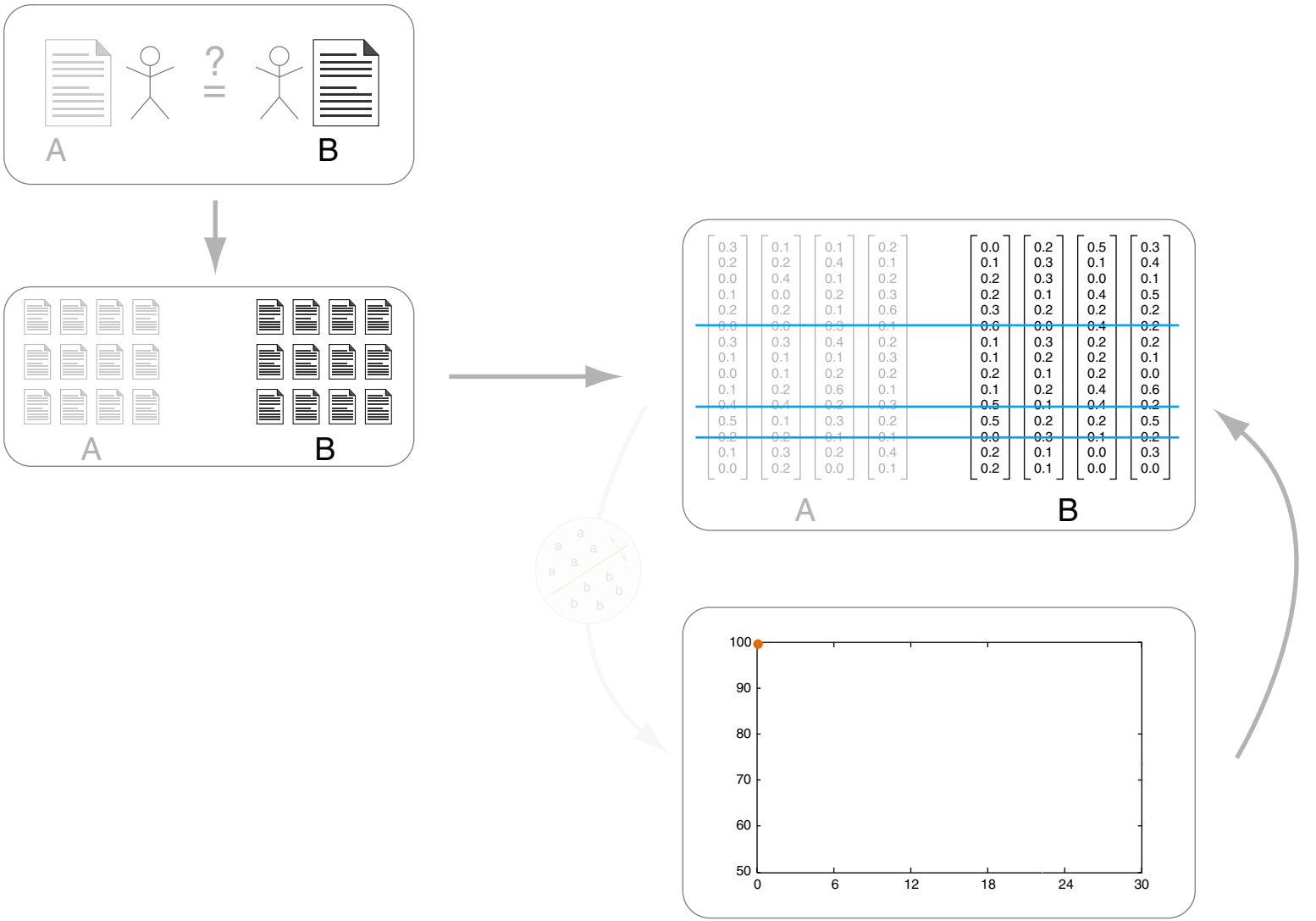
©Webis 2021

Authorship Verification via “Unmasking” [Koppel/Schler 2004]



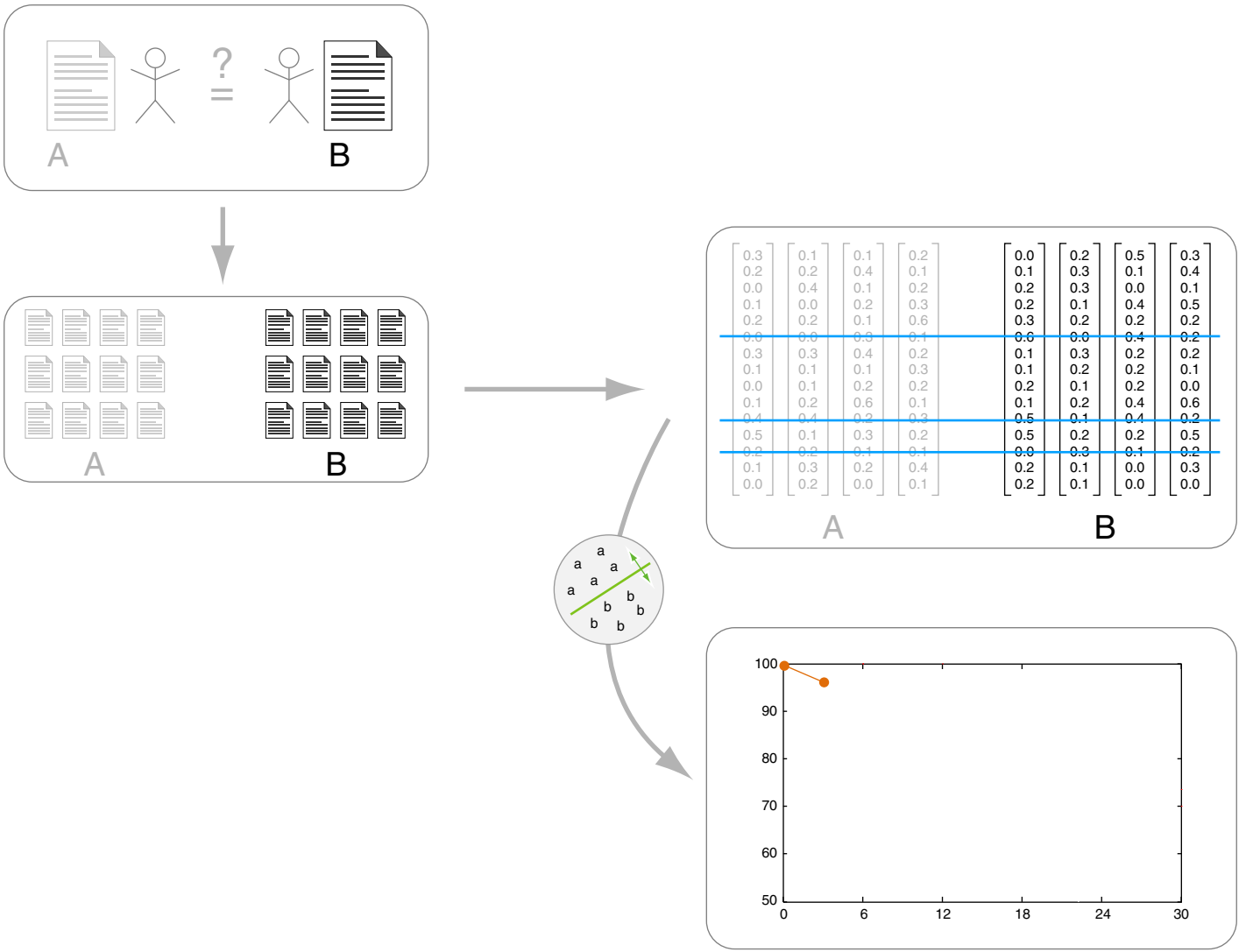
©Webis 2021

Authorship Verification via “Unmasking” [Koppel/Schler 2004]



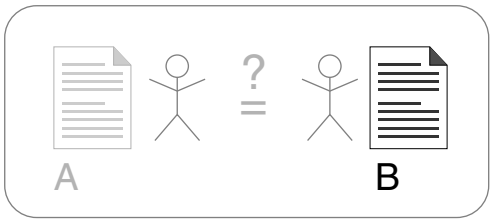
©Webis 2021

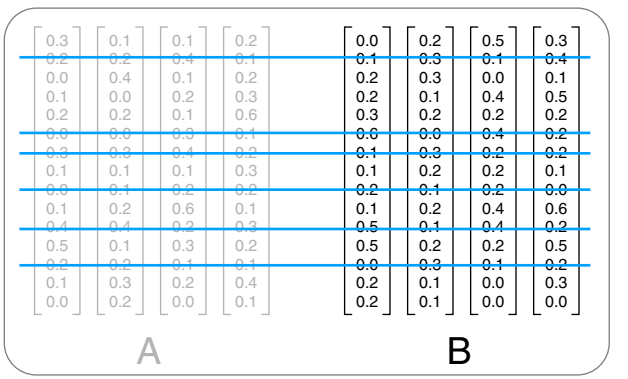
Authorship Verification via “Unmasking” [Koppel/Schler 2004]

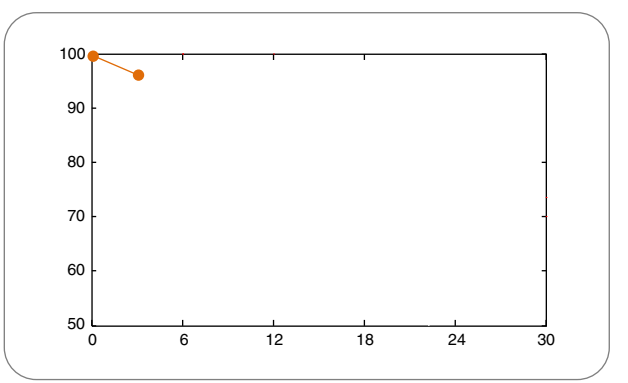


©Webis 2021

Authorship Verification via “Unmasking” [Koppel/Schler 2004]

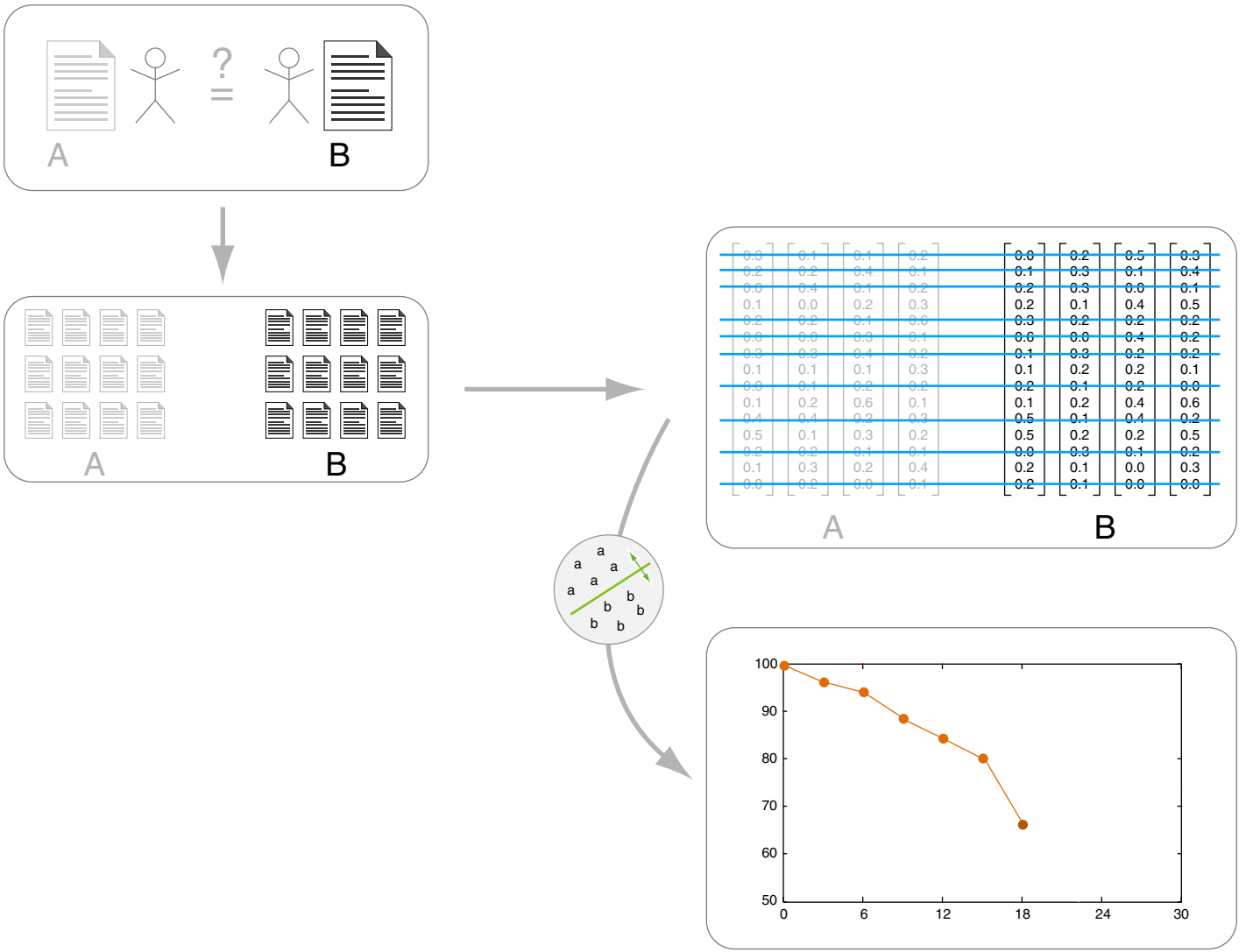






©Webis 2021

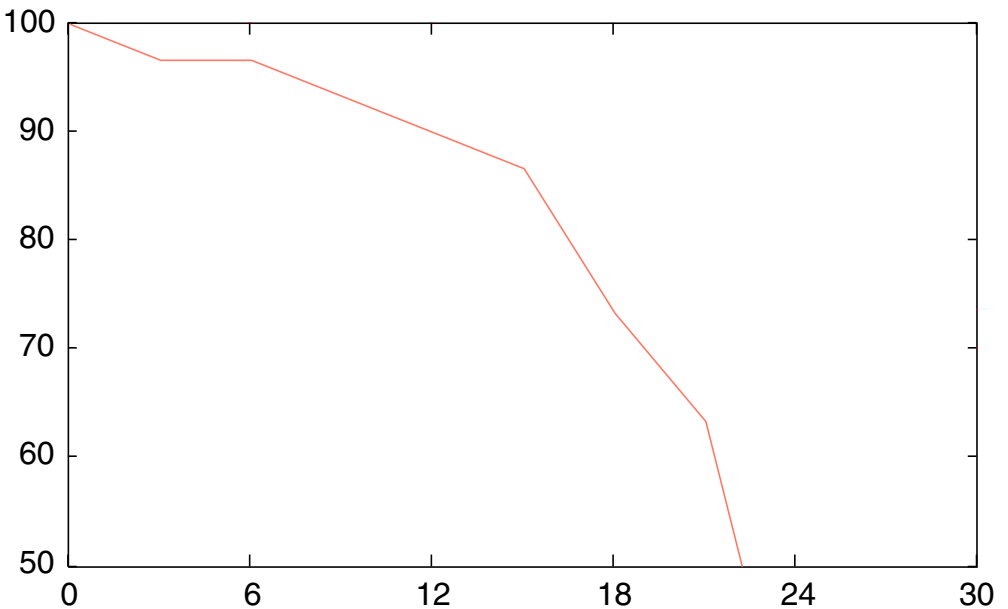
Authorship Verification via “Unmasking” [Koppel/Schler 2004]



©Webis 2021

Authorship Verification via “Unmasking”

Typical learning characteristic for . . .

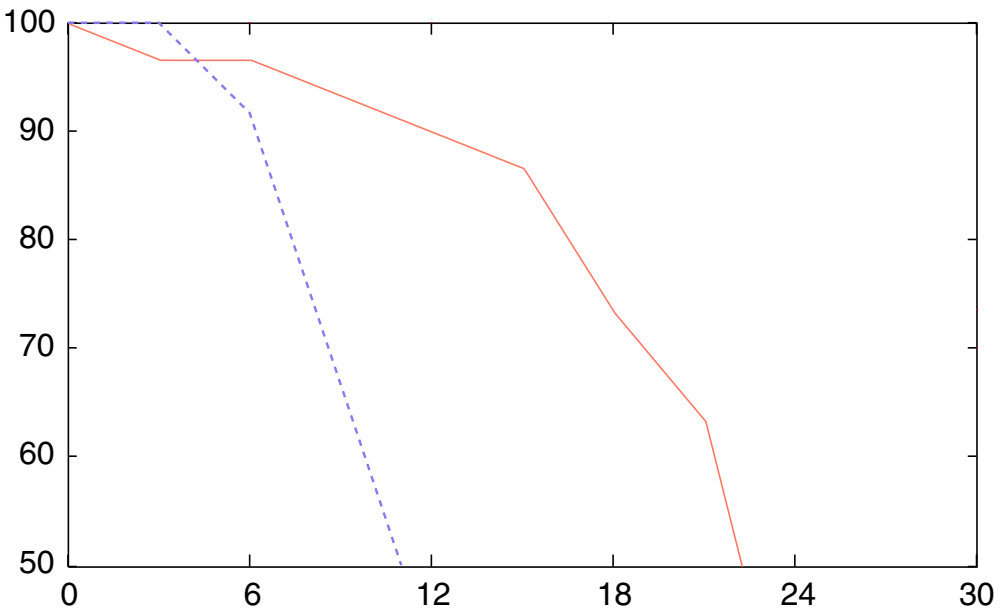


different authors ($A \neq B$)

©Webis 2021

Authorship Verification via “Unmasking”

Typical learning characteristic for . . .



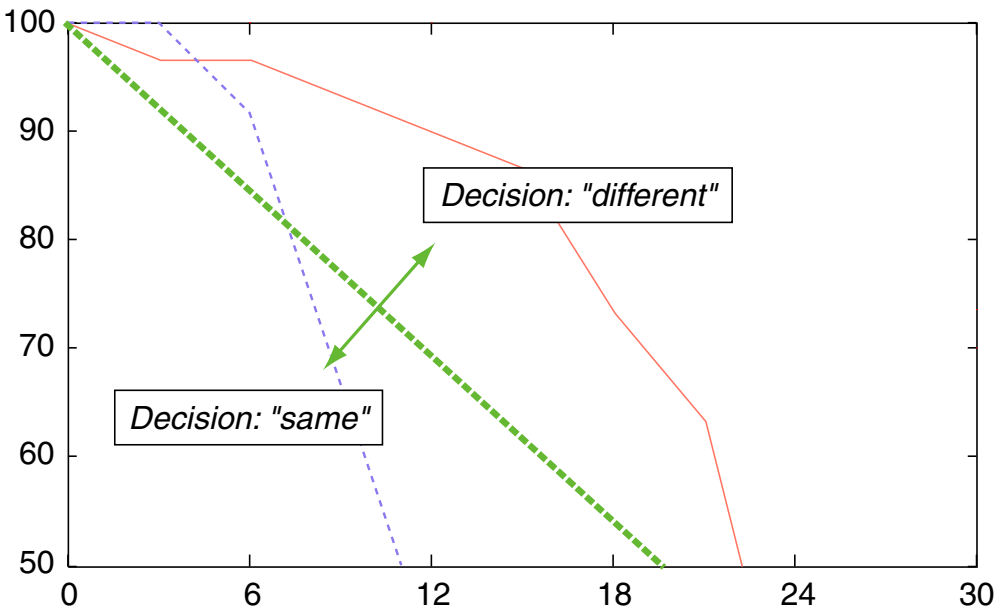
different authors $(A \neq B)$

same author $(A = B)$

©Webis 2021

Authorship Verification via “Unmasking”

Typical learning characteristic for . . .



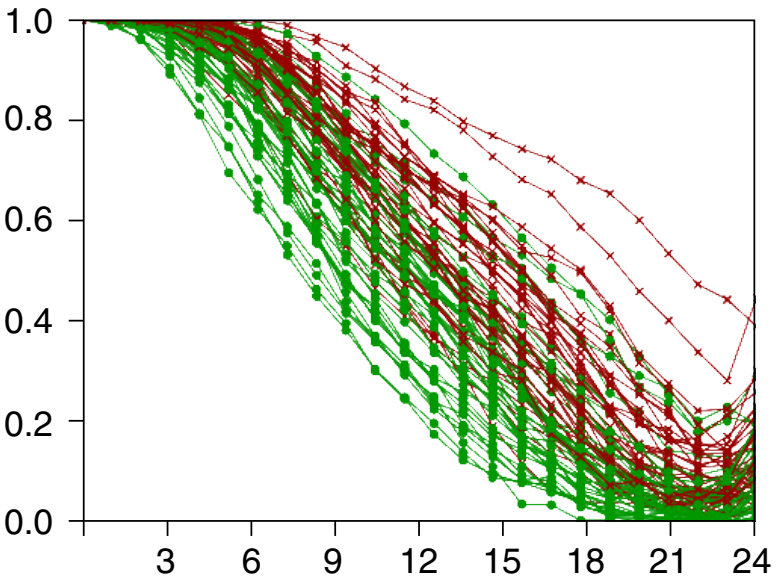
different authors ($A \neq B$)

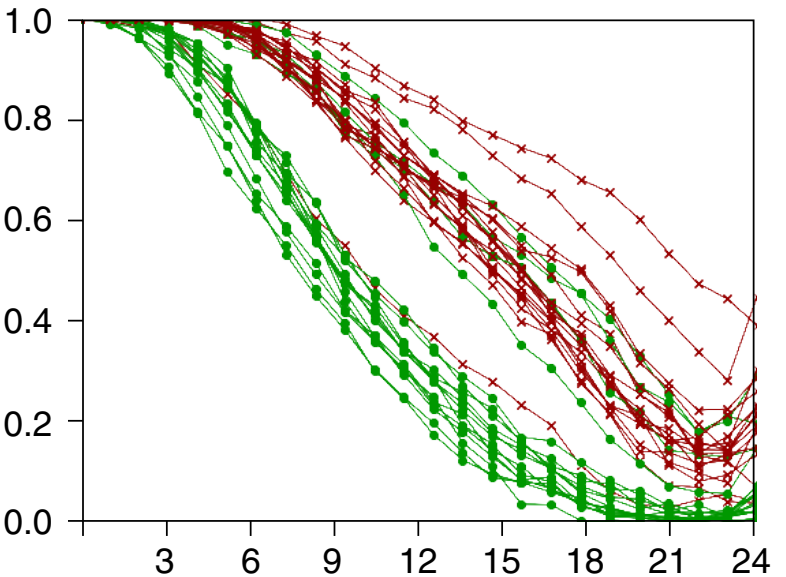
same author ($A = B$)

The typical learning characteristic can be learned.

©Webis 2021

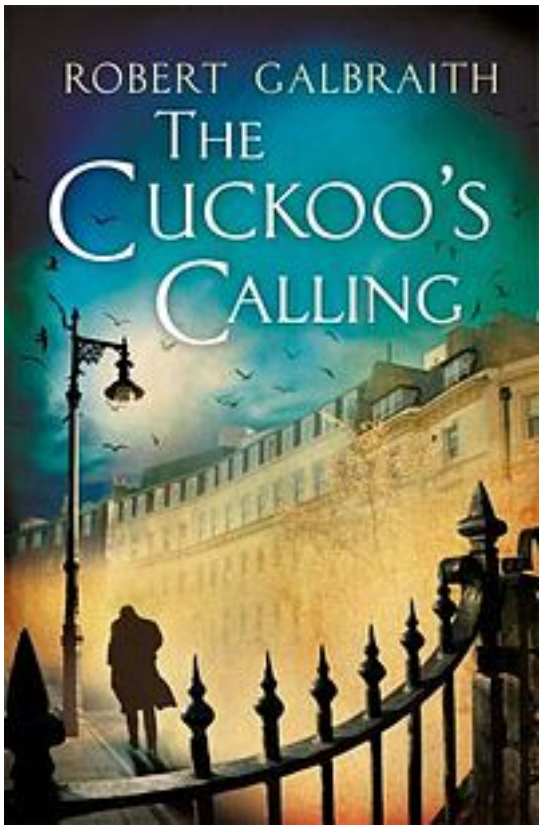
Authorship Verification via “Unmasking”





Applied to 78 pairs of texts, 4,000 words each ➜ 26% of decisions are “safe”

©Webis 2021



Fake likes

Fake news

Fake clicks

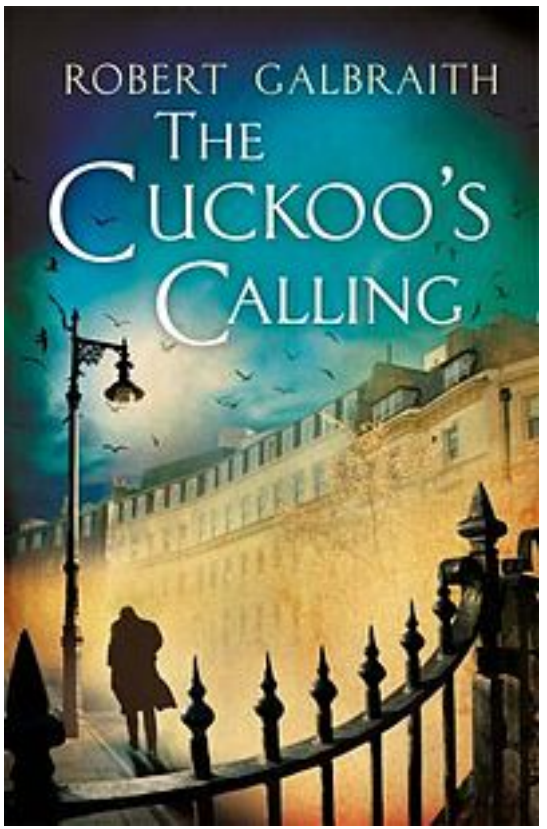
Fake users

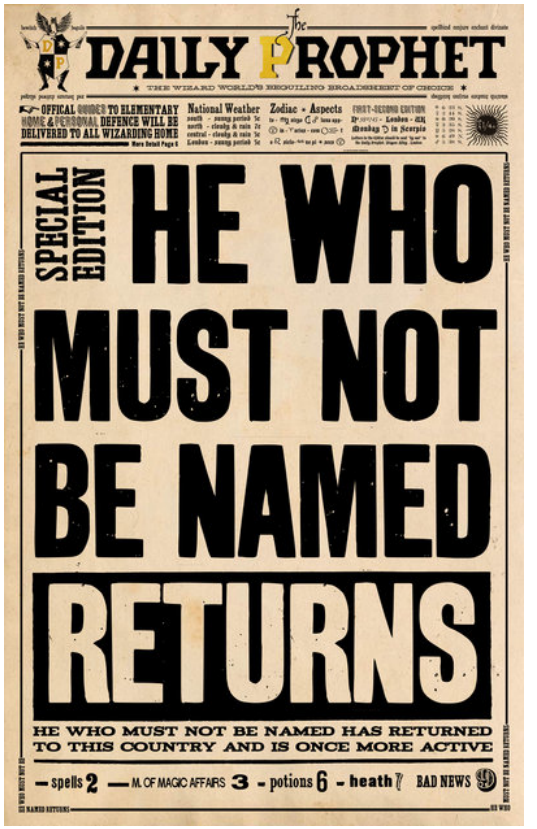
Fake reviews

Fake comments

⋮

Fake identities (pseudonyms)


ROBERT GALBRAITH
THE CUCKOO'S CALLING


The DAILY PROPHET
THE WIZARD WORLD'S BEGUILING BROADSHEET OF CHOICE
OFFICAL GUIDES TO ELEMENTARY HOME & PERSONAL DEFENCE WILL BE DELIVERED TO ALL WIZARDING HOME
National Weather
Zodiac • Aspects
FIRST-SECOND EDITION
SPECIAL EDITION
HE WHO MUST NOT BE NAMED RETURNS
HE WHO MUST NOT BE NAMED HAS RETURNED TO THIS COUNTRY AND IS ONCE MORE ACTIVE
– spells 2 — M. OF MAGIC AFFAIRS 3 - potions 6 - heath 7 BAD NEWS 9

Constrained Paraphrasing

To the Members of the California State Assembly:

I am returning Assembly Bill 1176 without my signature.

For some time now I have lamented the fact that major issues are overlooked while many unnecessary bills come to me for consideration. Water reform, prison reform, and health care are major issues my Administration has brought to the table, but the Legislature just kicks the can down the alley.

Yet another legislative year has come and gone without the major reforms Californians overwhelmingly deserve. In light of this, and after careful consideration, I believe it is unnecessary to sign this measure at this time.

Sincerely,

Arnold Schwarzenegger

[Veto message for the Shipyard project, Port of San Francisco. Oct. 12th, 2009]

©Webis 2021

To the Members of the California State Assembly:

I am returning Assembly Bill 1176 without my signature.

For some time now I have lamented the fact that major issues are overlooked while many unnecessary bills come to me for consideration. Water reform, prison reform, and health care are major issues my Administration has brought to the table, but the Legislature just kicks the can down the alley.

Yet another legislative year has come and gone without the major reforms Californians overwhelmingly deserve. In light of this, and after careful consideration, I believe it is unnecessary to sign this measure at this time.

Sincerely,

Arnold Schwarzenegger

[Veto message for the Shipyard project, Port of San Francisco. Oct. 12th, 2009]

©Webis 2021

To the Members of the California State Assembly:

I am returning Assembly Bill 1176 without my signature.

For some time now I have lamented the fact that major issues are overlooked while many unnecessary bills come to me for consideration. Water reform, prison reform, and health care are major issues my Administration has brought to the table, but the Legislature just kicks the can down the alley.

Yet another [illegible] Californians overwhelm[illegible] it is unnecess[illegible]

Sincerely,

Arnold Schwarzenegger

"*My goodness. What a coincidence* [. . .] "

[Aaron McLear, Schwarzenegger spokesman, Oct. 2009]

©Webis 2021

On Acrostics

An acrostic is a poem or other form of writing in which the first letter, syllable or word of each line, paragraph or other recurring feature in the text spells out a word or a message. [Wikipedia]

A poem [Kuperavage 2000] :

```
H  He broke my heart
E  Every piece, shattered
A  All I wanted was his love
R  Real, as he promised
T  True, as mine for him
...
```


©Webis 2021

On Acrostics

An acrostic is a poem or other form of writing in which the first letter, syllable or word of each line, paragraph or other recurring feature in the text spells out a word or a message. [Wikipedia]

A poem [Kuperavage 2000] :

```
H  He broke my heart
E  Every piece, shattered
A  All I wanted was his love
R  Real, as he promised
T  True, as mine for him
...
```

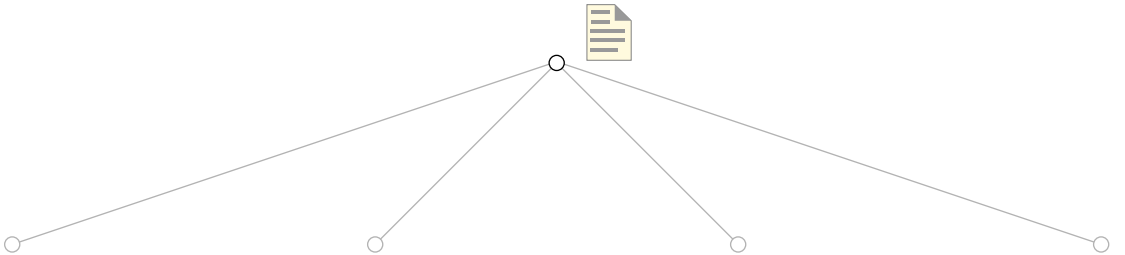
Task [Stein/Hagen/Bräutigam 2014]

Given: (1) A text T and an acrostic x.

(2) Lower and upper bounds on the desired line lengths.

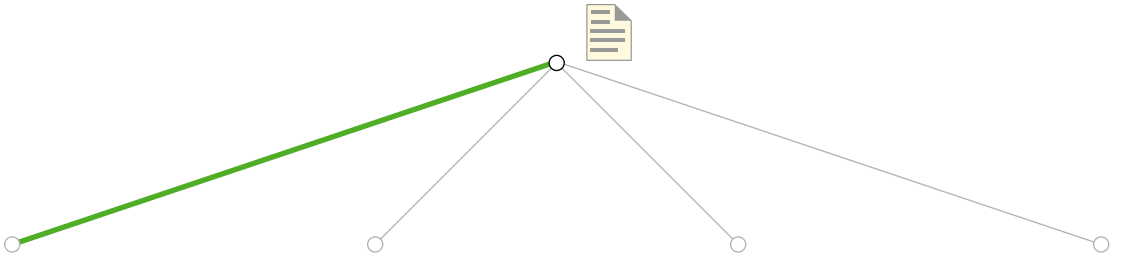
Task: Find a paraphrased version T^* of T in monospaced font that encodes x in some consecutive lines, if possible. Each line of T^* has to meet the length constraints.

©Webis 2021



Subtask: Create the character bauhaus

©Webis 2021

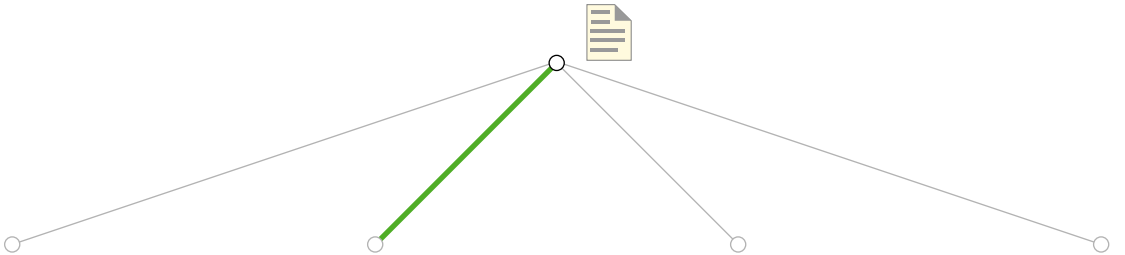


Before some time ~~now~~ I have lamented the fact that major issues are overlooked while many bills come to

«Preposition»

Subtask: Create the character **b**auhaus

©Webis 2021



Before some time now I have lamented the fact that major issues are overlooked while many bills come to

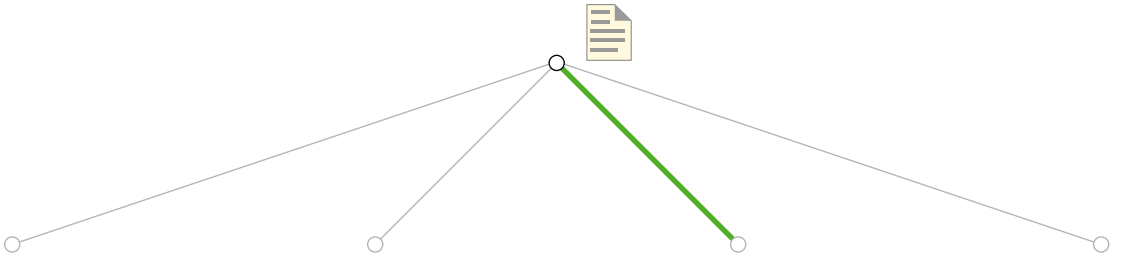
«Preposition»

For some time now I have lamented **but** the fact that major issues are overlooked while many bills

«Add Connective»

Subtask: Create the character **b**auhaus

©Webis 2021



Before some time ~~now~~ I have lamented the fact that major issues are overlooked while many bills come to

«Preposition»

For some time now I have lamented **but** the fact that major issues are overlooked while many bills

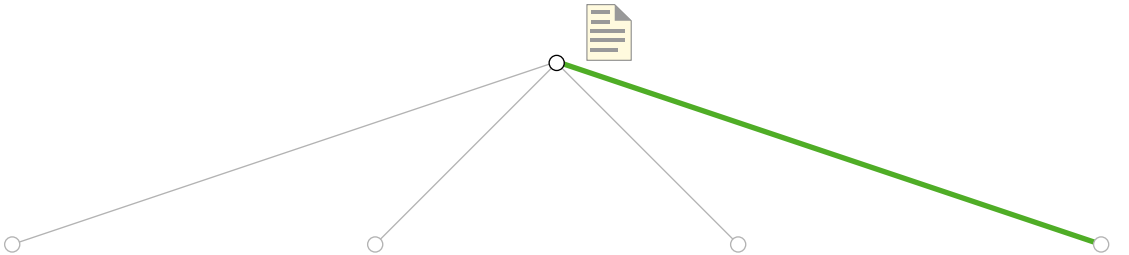
«Add Connective»

Been for some time now I have lamented the fact that major issues are overlooked while many bills come to

«Change Tense»

Subtask: Create the character **b**auhaus

©Webis 2021



Before some time ~~now~~ I have lamented the fact that major issues are overlooked while many bills come to

«Preposition»

For some time now I have lamented | **but** the fact that major issues are overlooked while many bills

«Add Connective»

Been for some time now I have lamented the fact that major issues are overlooked while many bills come to

«Change Tense»

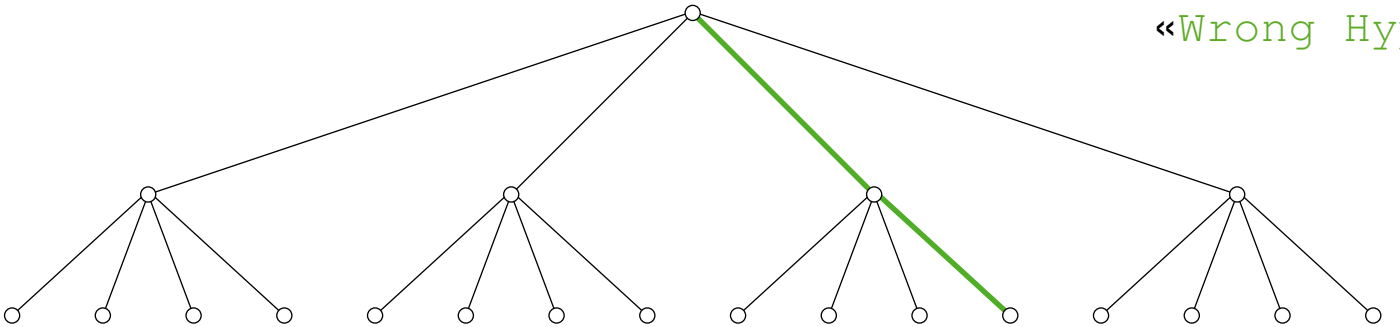
For some time now I have lamented the fact that major issues are overlooked while many |
bills come to

«Linebreak»

Subtask: Create the character **b**auhaus

©Webis 2021

«Wrong Hyphen»

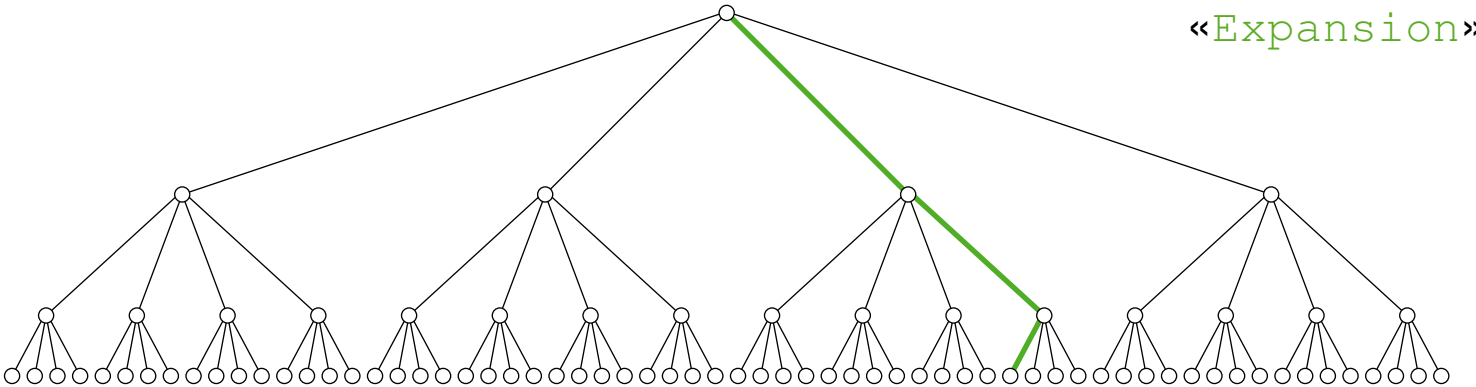


Subtask: Create the character bauhaus

©Webis 2021

Decision 3 of 64

«Expansion»

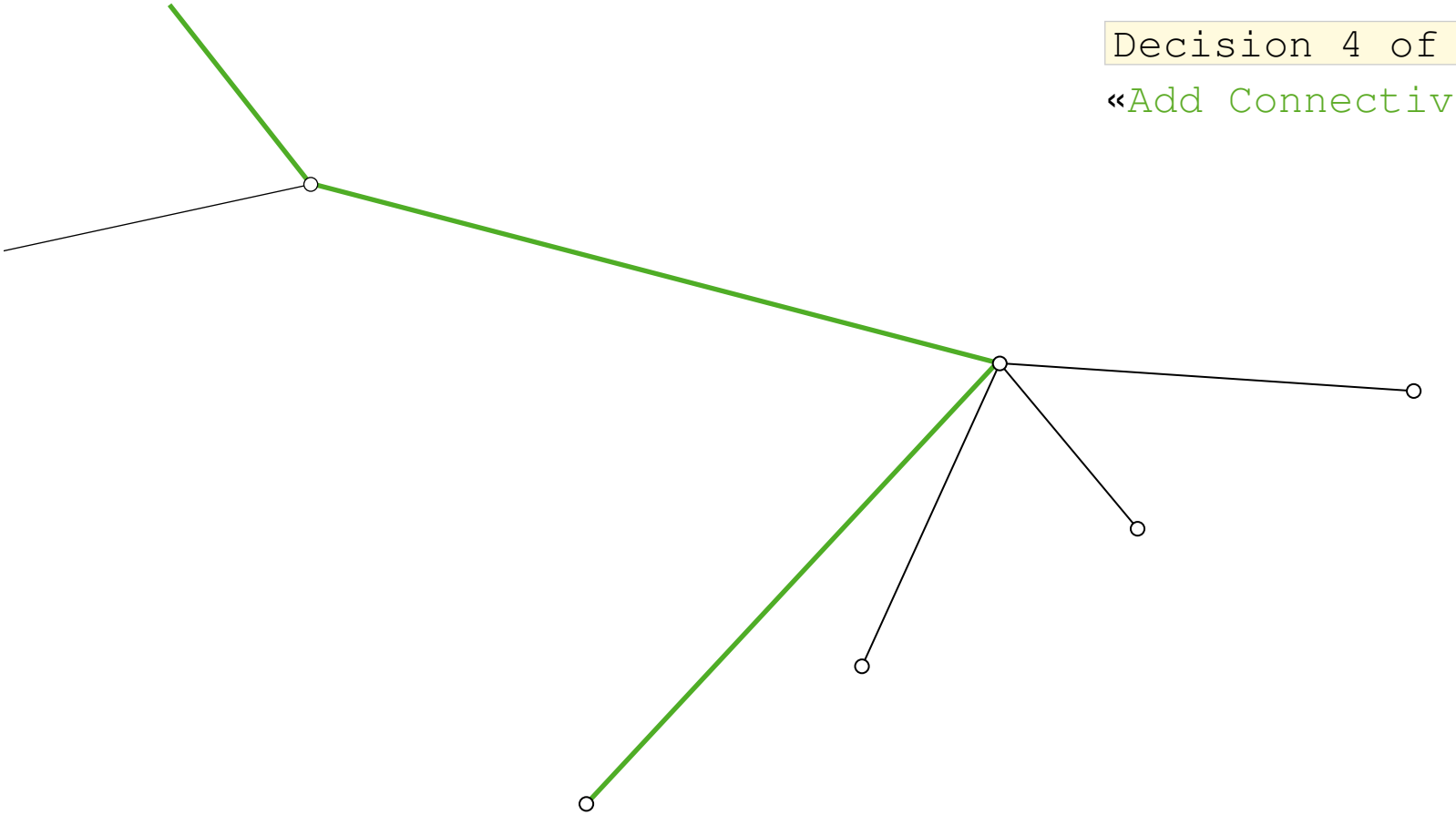


Subtask: Create the character ba**u**haus

©Webis 2021

Decision 4 of 256

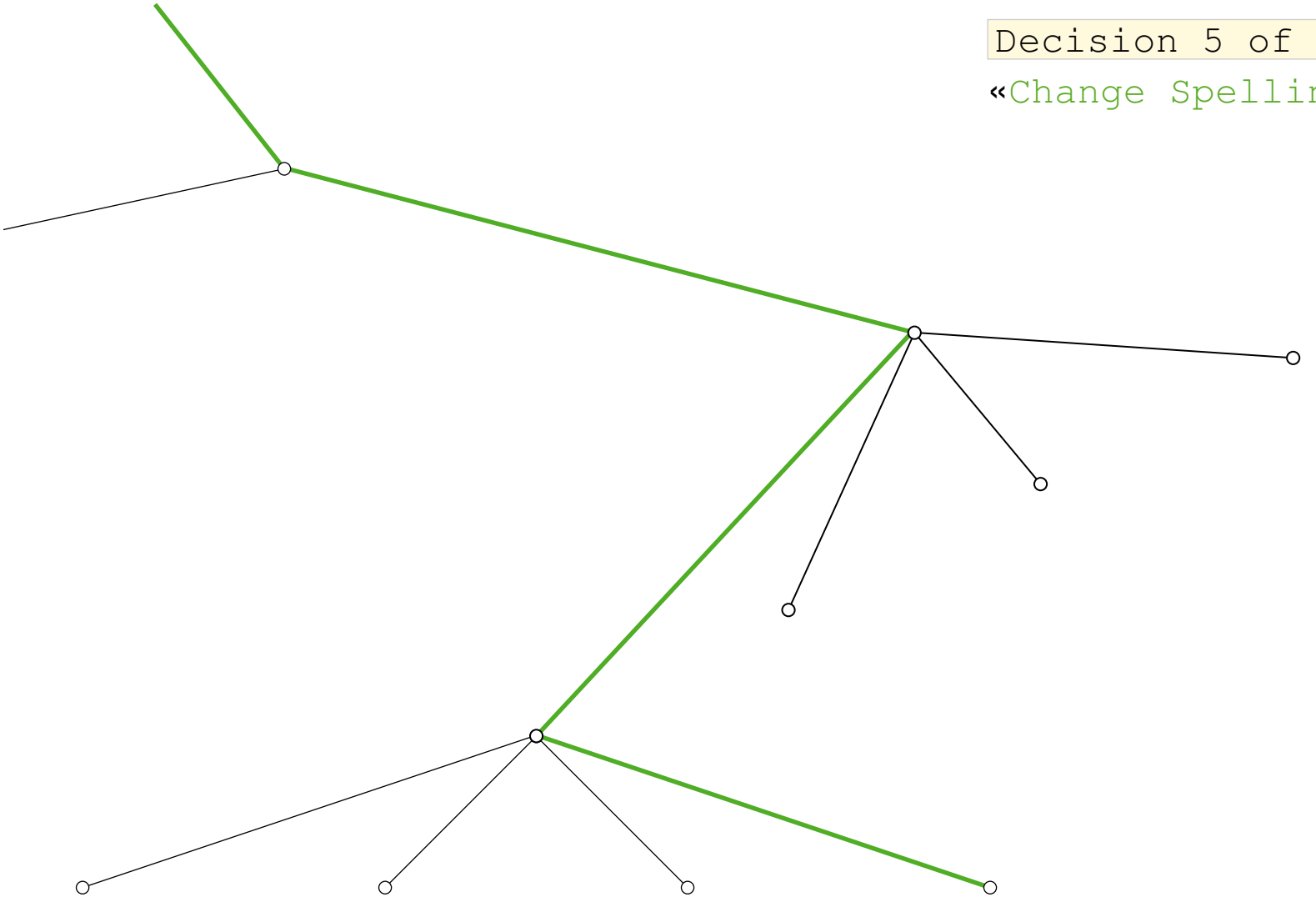
«Add Connective»



Subtask: Create the character bau**h**aus

©Webis 2021

«Change Spelling»

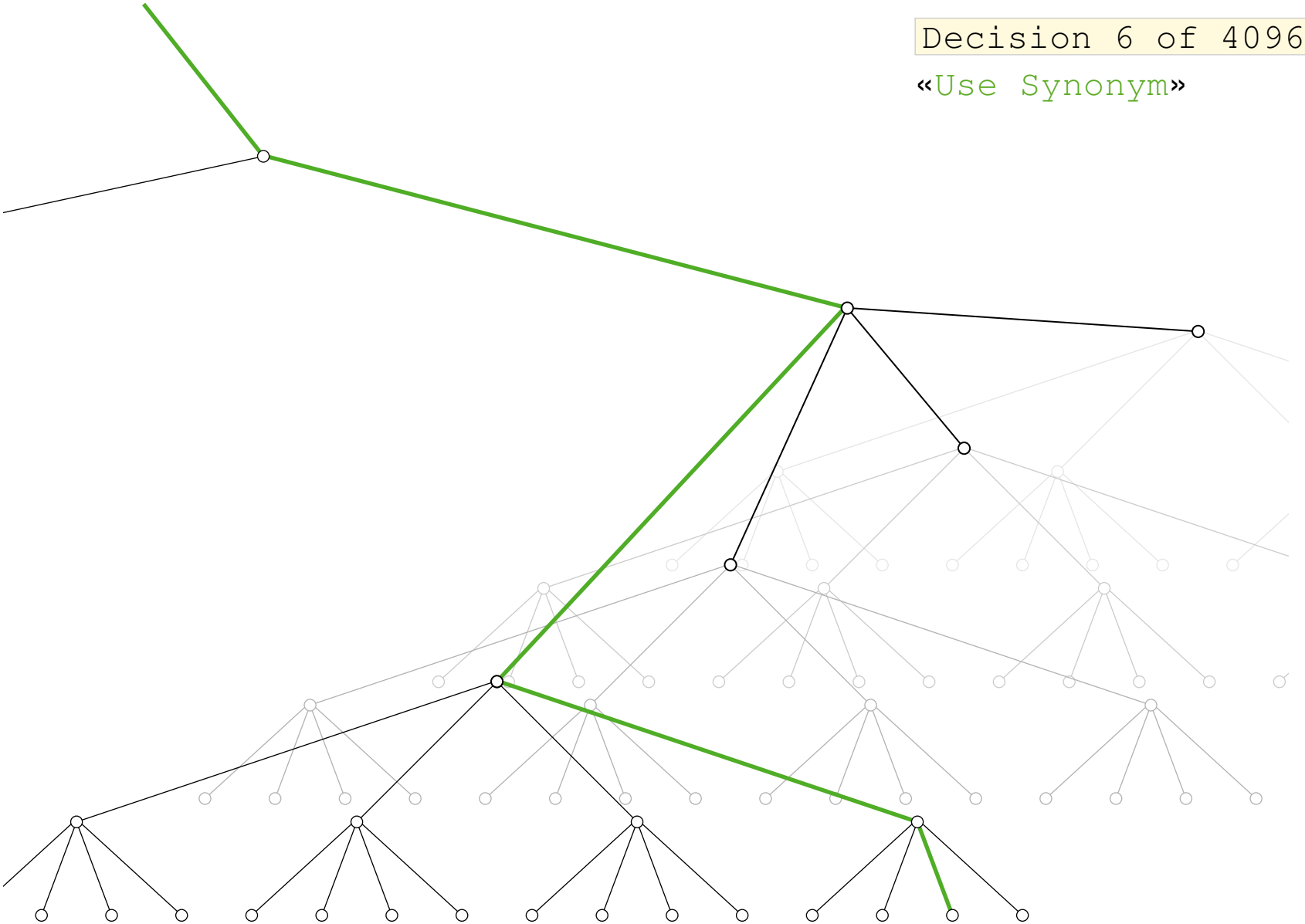


Subtask: Create the character bauh**a**us

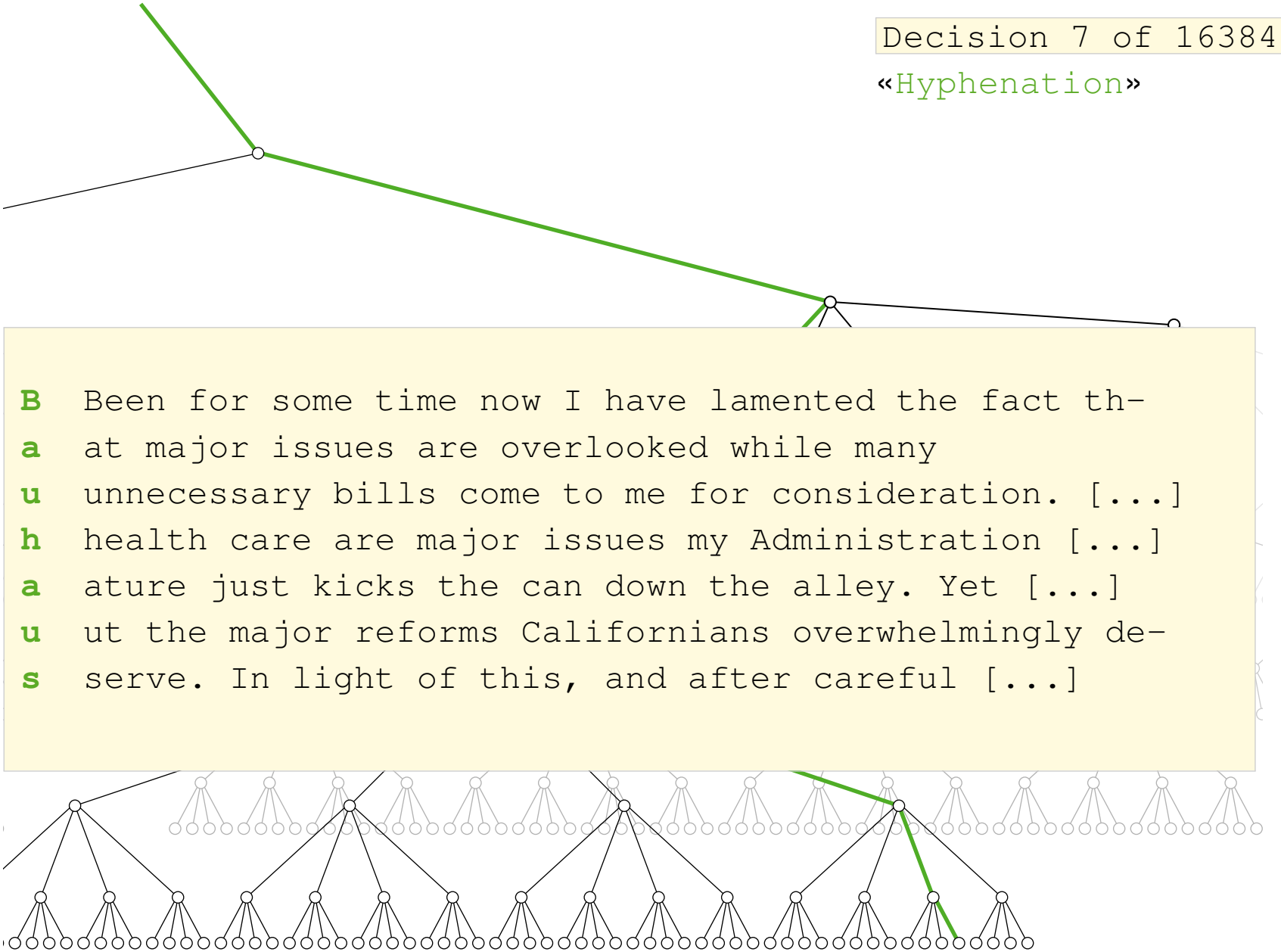
©Webis 2021

Decision 6 of 4096

«Use Synonym»



©Webis 2021



B Been for some time now I have lamented the fact th-
a at major issues are overlooked while many
u unnecessary bills come to me for consideration. [...]
h health care are major issues my Administration [...]
a ature just kicks the can down the alley. Yet [...]
u ut the major reforms Californians overwhelmingly de-
s serve. In light of this, and after careful [...]

©Webis 2021

Searchspace Facts

©Webis 2021

Searchspace Facts

Consider a text with a length of 100 words (the Schwarzenegger Letter) . . .

- $\approx$ 10 · 3 possibilities to change tense
- $\approx$ 100 possibilities to break a line
- $\approx$ 100 · 3 possibilities to introduce a synonym
- $\approx$ 100 · 3 possibilities to introduce filler words
- $\approx$ 100 · 5 possibilities to hyphenate a word
- $\gg$ 100 possibilities to introduce tautologies

  . . .

©Webis 2021

Searchspace Facts

Consider a text with a length of 100 words (the Schwarzenegger Letter) . . .

- $\approx$ $10 \cdot 3$ possibilities to change tense
- $\approx$ 100 possibilities to break a line
- $\approx$ $100 \cdot 3$ possibilities to introduce a synonym
- $\approx$ $100 \cdot 3$ possibilities to introduce filler words
- $\approx$ $100 \cdot 5$ possibilities to hyphenate a word
- $\gg$ 100 possibilities to introduce tautologies

  . . .

➜ $>$ 1 000 possible operations to generate a single letter of an acrostic

➜ $\mathcal{O}(10^{3n})$ possibilities to synthesize an $n = 7$ letter word like ‘`Bauhaus`’

Compare the following numbers:

10^{80} atoms in the observable universe

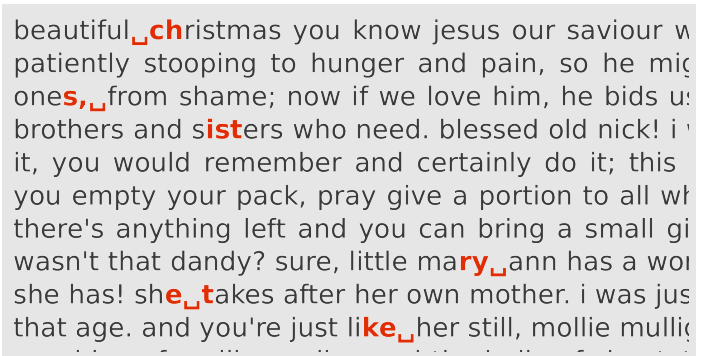
10^{123} game-tree complexity of chess

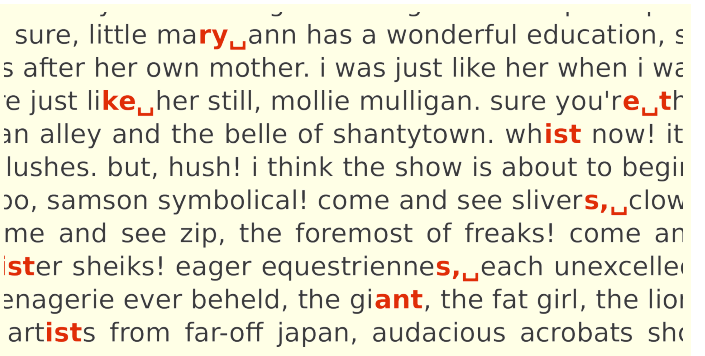
©Webis 2021

Toward Author Obfuscation [Stein/Potthast/Hagen/Bevendorff 2016]

©Webis 2021

Toward Author Obfuscation [Stein/Potthast/Hagen/Bevendorff 2016]



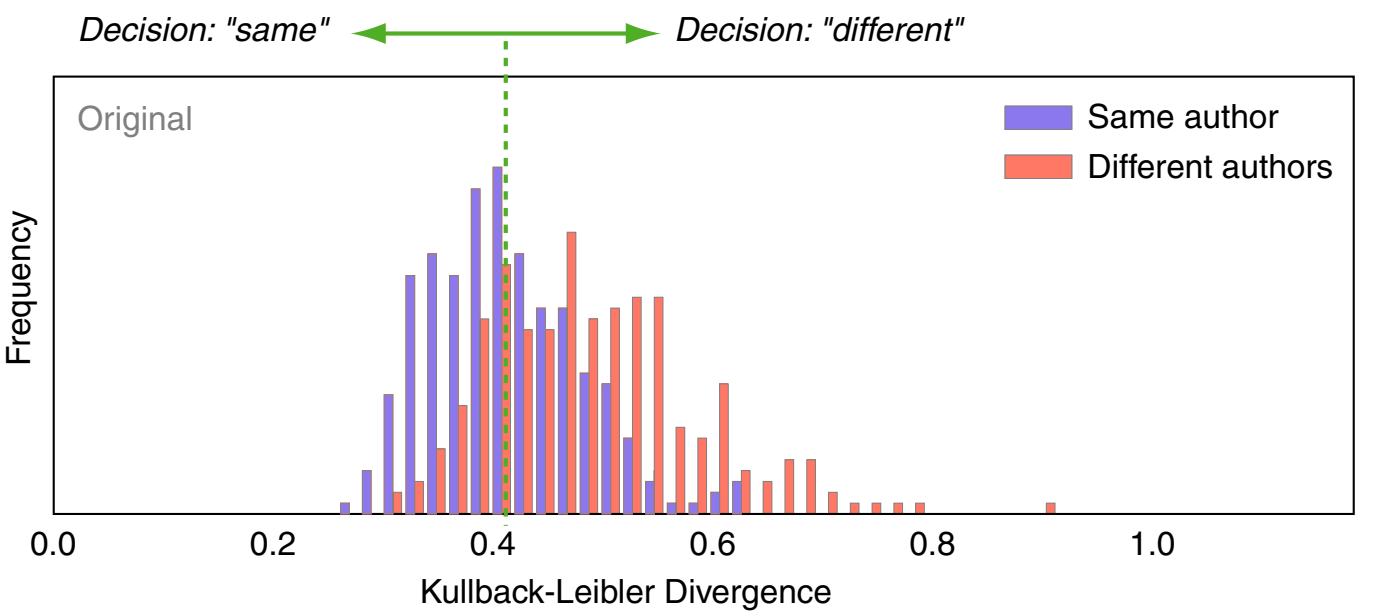


©Webis 2021

Toward Author Obfuscation [Stein/Potthast/Hagen/Bevendorff 2016]

beautiful␣**ch**ristmas you know jesus our saviour w
patiently stooping to hunger and pain, so he mig
one**s,**␣from shame; now if we love him, he bids u
brothers and s**ist**ers who need. blessed old nick! i
it, you would remember and certainly do it; this
you empty your pack, pray give a portion to all wh
there's anything left and you can bring a small gi
wasn't that dandy? sure, little ma**ry**␣ann has a wo
she has! sh**e**␣**t**akes after her own mother. i was jus
that age. and you're just li**ke**␣her still, mollie mulli

sure, little ma**ry**␣ann has a wonderful education, s
s after her own mother. i was just like her when i wa
e just li**ke**␣her still, mollie mulligan. sure you're**e**␣**t**h
an alley and the belle of shantytown. wh**ist** now! it
lushes. but, hush! i think the show is about to begi
oo, samson symbolical! come and see sliver**s,**␣clow
me and see zip, the foremost of freaks! come an
ister sheiks! eager equestrienne**s,**␣each unexcelle
enagerie ever beheld, the gi**ant**, the fat girl, the lio
art**ist**s from far-off japan, audacious acrobats sh

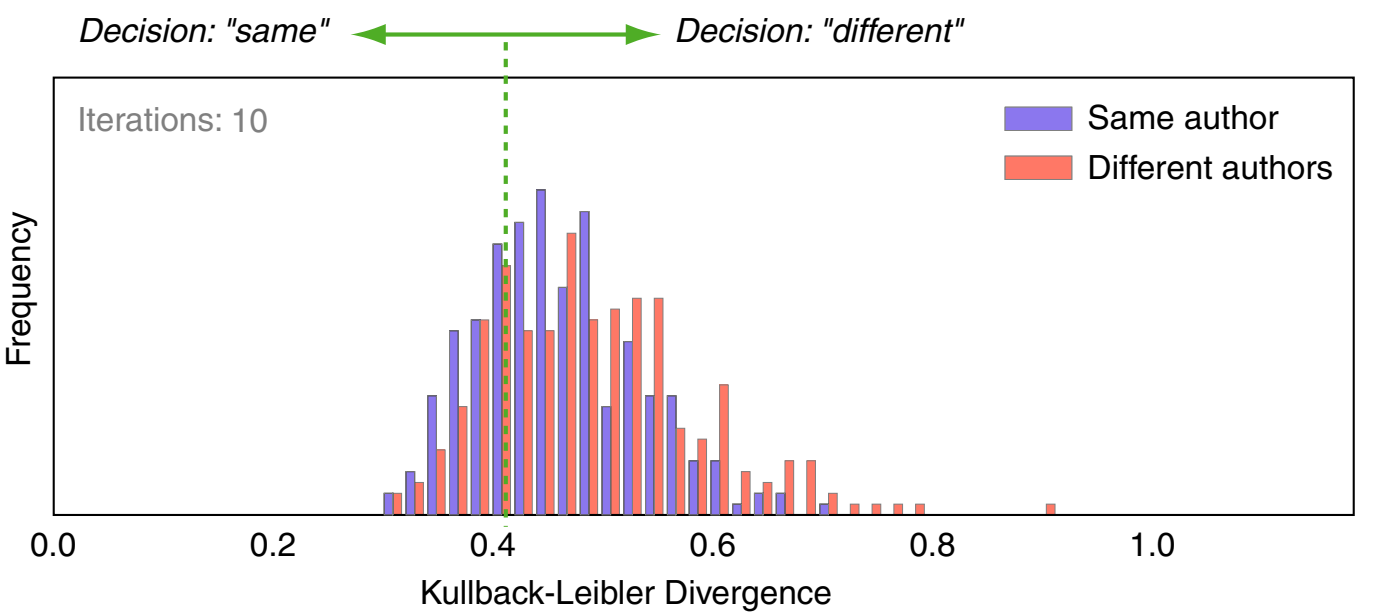


©Webis 2021

Toward Author Obfuscation [Stein/Potthast/Hagen/Bevendorff 2016]

beautiful␣**ch**ristmas you know jesus our saviour w
patiently stooping to hunger and pain, so he mig
one**s,**␣from shame; now if we love him, he bids us
brothers and s**ist**ers who need. blessed old nick! i
it, you would remember and certainly do it; this
you empty your pack, pray give a portion to all wh
there's anything left and you can bring a small gi
wasn't that dandy? sure, little ma**ry**␣ann has a wo
she has! sh**e**␣**t**akes after her own mother. i was jus
that age. and you're just li**ke**␣her still, mollie mulli

sure, little ma**ry**␣ann has a wonderful education, s
s after her own mother. i was just like her when i wa
e just li**ke**␣her still, mollie mulligan. sure you're**e**␣**t**h
an alley and the belle of shantytown. wh**ist** now! it
lushes. but, hush! i think the show is about to begi
oo, samson symbolical! come and see sliver**s,**␣clow
me and see zip, the foremost of freaks! come an
ister sheiks! eager equestrienne**s,**␣each unexcelle
enagerie ever beheld, the gi**ant**, the fat girl, the lio
arti**st**s from far-off japan, audacious acrobats sh

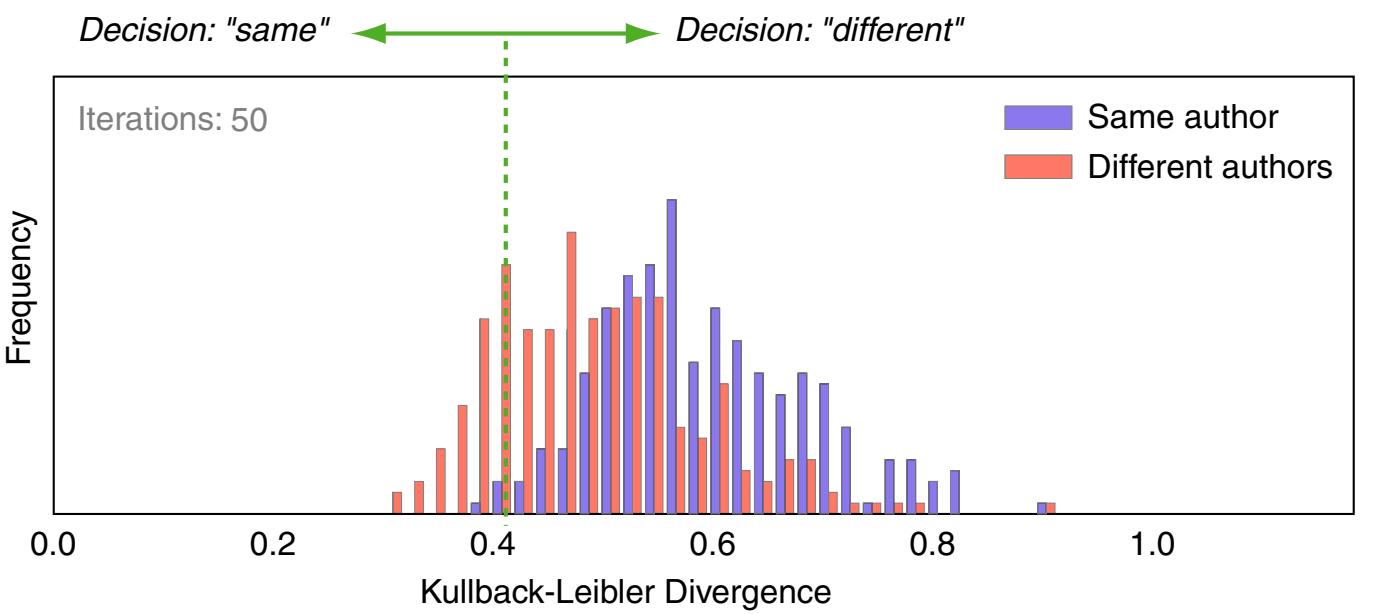


©Webis 2021

Toward Author Obfuscation [Stein/Potthast/Hagen/Bevendorff 2016]

beautiful␣**ch**ristmas you know jesus our saviour w
patiently stooping to hunger and pain, so he mig
one**s,**␣from shame; now if we love him, he bids us
brothers and s**ist**ers who need. blessed old nick! i
it, you would remember and certainly do it; this
you empty your pack, pray give a portion to all wh
there's anything left and you can bring a small gi
wasn't that dandy? sure, little ma**ry**␣ann has a wo
she has! sh**e**␣**t**akes after her own mother. i was jus
that age. and you're just li**ke**␣her still, mollie mulli

sure, little ma**ry**␣ann has a wonderful education, s
s after her own mother. i was just like her when i wa
e just li**ke**␣her still, mollie mulligan. sure you're**e**␣**t**h
an alley and the belle of shantytown. wh**ist** now! it
lushes. but, hush! i think the show is about to begi
oo, samson symbolical! come and see sliver**s,**␣clow
me and see zip, the foremost of freaks! come an
ister sheiks! eager equestrienne**s,**␣each unexcelle
enagerie ever beheld, the gi**ant**, the fat girl, the lio
art**ist**s from far-off japan, audacious acrobats sh

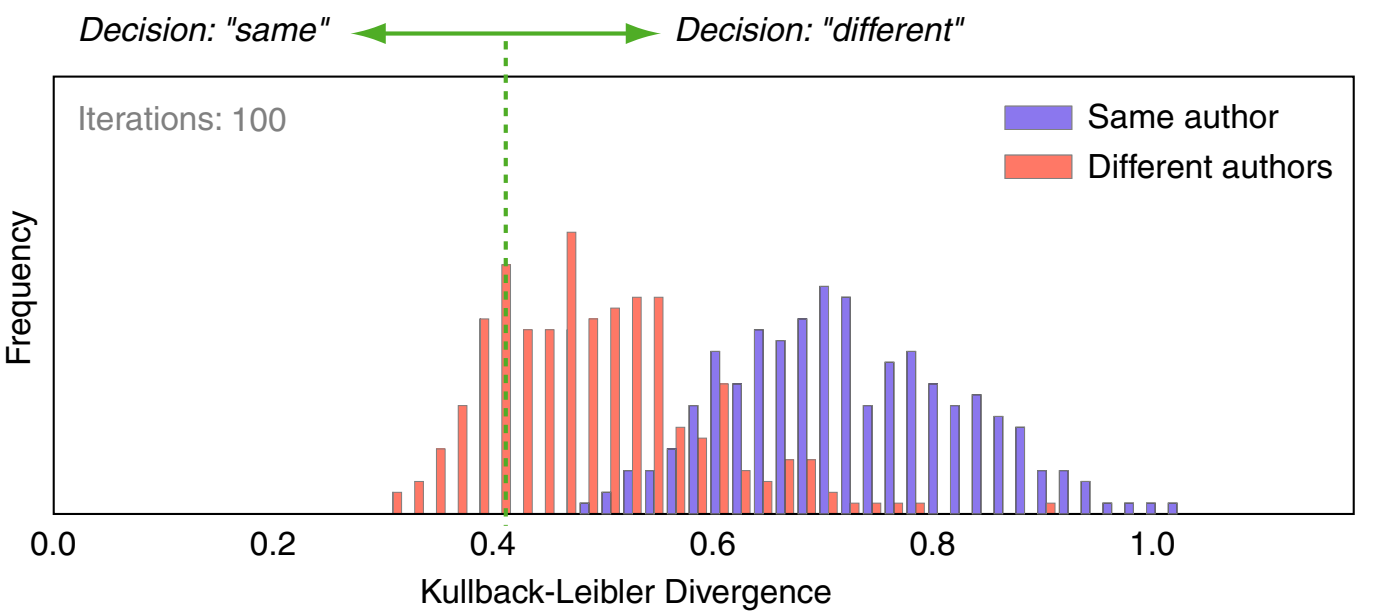


©Webis 2021

Toward Author Obfuscation [Stein/Potthast/Hagen/Bevendorff 2016]

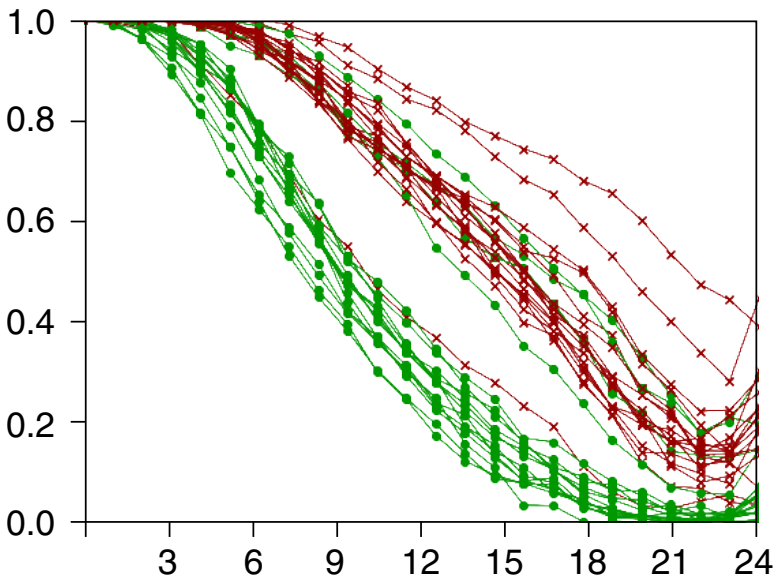
beautiful␣**ch**ristmas you know jesus our saviour w
patiently stooping to hunger and pain, so he mig
one**s,**␣from shame; now if we love him, he bids u
brothers and s**ist**ers who need. blessed old nick! i
it, you would remember and certainly do it; this
you empty your pack, pray give a portion to all wh
there's anything left and you can bring a small gi
wasn't that dandy? sure, little ma**ry**␣ann has a wo
she has! sh**e**␣**t**akes after her own mother. i was jus
that age. and you're just li**ke**␣her still, mollie mulli

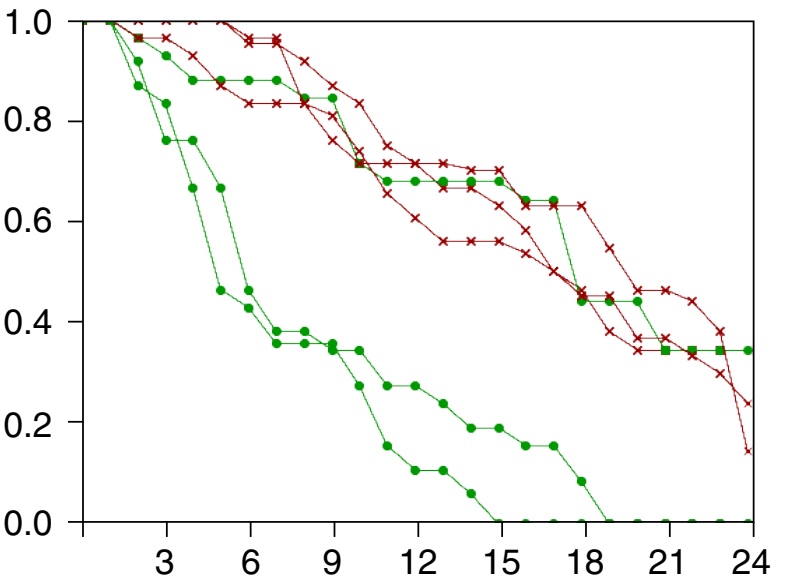
sure, little ma**ry**␣ann has a wonderful education, s
s after her own mother. i was just like her when i wa
e just li**ke**␣her still, mollie mulligan. sure you're**e**␣**t**h
an alley and the belle of shantytown. wh**ist** now! it
lushes. but, hush! i think the show is about to begi
oo, samson symbolical! come and see sliver**s,**␣clow
me and see zip, the foremost of freaks! come an
ister sheiks! eager equestrienne**s,**␣each unexcelle
enagerie ever beheld, the gi**ant**, the fat girl, the lio
art**ist**s from far-off japan, audacious acrobats sh



©Webis 2021

Authorship Verification via “Unmasking”





26% of decisions are “safe” ➜ 10% remain safe when obfuscating 1% of the text

©Webis 2021

Conclusion

Summary

- Constrained paraphrasing via heuristic search in style space
- 5 author obfuscators vs. 44 authorship verifiers in 4 settings
- Authorship verifiers represent the state of the art as per PAN'13/14/15
- Obfuscators flip on average from 20% up to 49% of true positive decisions

©Webis 2021

Conclusion

Summary

- Constrained paraphrasing via heuristic search in style space
- 5 author obfuscators vs. 44 authorship verifiers in 4 settings
- Authorship verifiers represent the state of the art as per PAN'13/14/15
- Obfuscators flip on average from 20% up to 49% of true positive decisions

Take-away messages

- State of the art in authorship verification vulnerable to obfuscation
- Automatic obfuscation is feasible, yet far from perfection
- Hardly anyone considers obfuscation a threat

©Webis 2021

Conclusion

Summary

- Constrained paraphrasing via heuristic search in style space
- 5 author obfuscators vs. 44 authorship verifiers in 4 settings
- Authorship verifiers represent the state of the art as per PAN'13/14/15
- Obfuscators flip on average from 20% up to 49% of true positive decisions

Take-away messages

- State of the art in authorship verification vulnerable to obfuscation
- Automatic obfuscation is feasible, yet far from perfection
- Hardly anyone considers obfuscation a threat

- Author obfuscation and author identification are locked in an instance of the "Potter-Voldemort Conundrum":

Neither can live while the other survives

©Webis 2021

Conclusion

Summary

- Constrained paraphrasing via heuristic search in style space
- 5 author obfuscators vs. 44 authorship verifiers in 4 settings
- Authorship verifiers represent the state of the art as per PAN'13/14/15
- Obfuscators flip on average from 20% up to 49% of true positive decisions

Take-away messages

- State of the art in authorship verification vulnerable to obfuscation
- Automatic obfuscation is feasible, yet far from perfection
- Hardly anyone considers obfuscation a threat

- Author obfuscation and author identification are locked in an instance of the "Potter-Voldemort Conundrum":

Neither can live while the other survives

Thank you for your attention!

©Webis 2021