

Web Page Segmentation Revisited: Evaluation Framework and Dataset

CIKM 2020



Johannes
Kiesel¹



Florian
Kneist¹



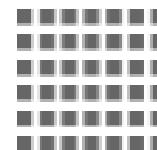
Lars
Meyer¹



Kristof
Komlossy¹



Benno
Stein¹



Web Page Segmentation

Flashback: Supercut of Elton John singing 'Your Song' through the years

posted by Samantha Martin | Popdust - 4 years ago

comment share

Listen to Elton John on iHeartRadio

There's something exciting about being among the first few thousand people to watch a YouTube video on the verge of virality. Congrats! You're about to be one of those people (assuming you read this close to the publish date).

Elton John's "Your Song" is such a perfect marriage of singer to song. It's no wonder that it's one of the most popular wedding songs of all time. In a slew of major hits, "Your Song" manages to be John's signature song.

This soon-to-be viral video is a supercut of John singing the song through the years since it was first released in 1970. While his performance is miraculously consistent, his wardrobe is anything but. (What exactly was that Donald Duck costume about?) So sit on the rooftop, kick off the moss, and enjoy.

And you can tell everybody...you saw it first.

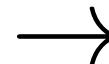


Photo Credit Getty Images



Comments

[View Comments](#)



Flashback: Supercut of Elton John singing 'Your Song' through the years

posted by Samantha Martin | Popdust - 4 years ago

comment share

Listen to Elton John on iHeartRadio

There's something exciting about being among the first few thousand people to watch a YouTube video on the verge of virality. Congrats! You're about to be one of those people (assuming you read this close to the publish date).

Elton John's "Your Song" is such a perfect marriage of singer to song. It's no wonder that it's one of the most popular wedding songs of all time. In a slew of major hits, "Your Song" manages to be John's signature song.

This soon-to-be viral video is a supercut of John singing the song through the years since it was first released in 1970. While his performance is miraculously consistent, his wardrobe is anything but. (What exactly was that Donald Duck costume about?) So sit on the rooftop, kick off the moss, and enjoy.

And you can tell everybody...you saw it first.



Photo Credit Getty Images



Comments

[View Comments](#)

[Terms of Use](#) [Privacy Statement](#) [Copyright and Trademark Notice](#) [EOC Public File](#) [Public Inspection File](#) [iHeartRadio - Get the Free Radio App](#) [General Contest Guidelines](#)
Advertise on News Radio 610 WTVN Get Results Contact Philanthropy
© 2017 iHeartMedia, Inc.



Listen Now on

[Terms of Use](#) [Privacy Statement](#) [Copyright and Trademark Notice](#) [EOC Public File](#) [Public Inspection File](#) [iHeartRadio - Get the Free Radio App](#) [General Contest Guidelines](#)
Advertise on News Radio 610 WTVN Get Results Contact Philanthropy
© 2017 iHeartMedia, Inc.



Listen Now on

Web Page Segmentation

610 WTVN Listen Live on iHeartRADIO News Radio 610 WTVN - News, Traffic, & Weather - Columbus, OH On-Air News Podcasts & Media Connect Contests

Search icon

Flashback: Supercut of Elton John singing 'Your Song' through the years

posted by Samantha Martin | Popdust - 4 years ago

comment share

Listen to Elton John on iHeartRadio

There's something exciting about being among the first few thousand people to watch a YouTube video on the verge of virality. Congrats! You're about to be one of those people (assuming you read this close to the publish date).

Elton John's "Your Song" is such a perfect marriage of singer to song. It's no wonder that it's one of the most popular wedding songs of all time. In a slew of major hits, "Your Song" manages to be John's signature song.

This soon-to-be viral video is a supercut of John singing the song through the years since it was first released in 1970. While his performance is miraculously consistent, his wardrobe is anything but. (What exactly was that Donald Duck costume about?) So sit on the rooftop, kick off the moss, and enjoy.

And you can tell everybody...you saw it first.



Photo Credit Getty Images



Comments

[View Comments](#)

610 WTVN Listen Live on iHeartRADIO News Radio 610 WTVN - News, Traffic, & Weather - Columbus, OH On-Air News Podcasts & Media Connect Contests

Search icon

Flashback: Supercut of Elton John singing 'Your Song' through the years

posted by Samantha Martin | Popdust - 4 years ago

comment share

Listen to Elton John on iHeartRadio

There's something exciting about being among the first few thousand people to watch a YouTube video on the verge of virality. Congrats! You're about to be one of those people (assuming you read this close to the publish date).

Elton John's "Your Song" is such a perfect marriage of singer to song. It's no wonder that it's one of the most popular wedding songs of all time. In a slew of major hits, "Your Song" manages to be John's signature song.

This soon-to-be viral video is a supercut of John singing the song through the years since it was first released in 1970. While his performance is miraculously consistent, his wardrobe is anything but. (What exactly was that Donald Duck costume about?) So sit on the rooftop, kick off the moss, and enjoy.

And you can tell everybody...you saw it first.



Photo Credit Getty Images



Comments

[View Comments](#)

Terms of Use Privacy Statement Copyright and Trademark Notice EEO Public File Public Inspection File iHeartRadio - Get the Free Radio App General Contest Guidelines Advertise on News Radio 610 WTVN Get Results Contact Philanthropy © 2017 iHeartMedia, Inc.

Listen Now on iHeartRADIO

Terms of Use Privacy Statement Copyright and Trademark Notice EEO Public File Public Inspection File iHeartRadio - Get the Free Radio App General Contest Guidelines Advertise on News Radio 610 WTVN Get Results Contact Philanthropy © 2017 iHeartMedia, Inc.

Listen Now on iHeartRADIO

Web Page Segmentation

610 WTVN Listen Live on iHeartRADIO News Radio 610 WTVN - News, Traffic, & Weather - Columbus, OH On-Air News Podcasts & Media Connect Contests

Flashback: Supercut of Elton John singing 'Your Song' through the years

posted by Samantha Martin | Popdust - 4 years ago

comment share

Listen to Elton John on iHeartRadio

There's something exciting about being among the first few thousand people to watch a YouTube video on the verge of virality. Congrats! You're about to be one of those people (assuming you read this close to the publish date).

Elton John's "Your Song" is such a perfect marriage of singer to song. It's no wonder that it's one of the most popular wedding songs of all time. In a slew of major hits, "Your Song" manages to be John's signature song.

This soon-to-be viral video is a supercut of John singing the song through the years since it was first released in 1970. While his performance is miraculously consistent, his wardrobe is anything but. (What exactly was that Donald Duck costume about?) So sit on the rooftop, kick off the moss, and enjoy.

And you can tell everybody...you saw it first.



Photo Credit Getty Images



Comments

View Comments



Flashback: Supercut of Elton John singing 'Your Song' through the years

posted by Samantha Martin | Popdust - 4 years ago

comment share

Listen to Elton John on iHeartRadio

There's something exciting about being among the first few thousand people to watch a YouTube video on the verge of virality. Congrats! You're about to be one of those people (assuming you read this close to the publish date).

Elton John's "Your Song" is such a perfect marriage of singer to song. It's no wonder that it's one of the most popular wedding songs of all time. In a slew of major hits, "Your Song" manages to be John's signature song.

This soon-to-be viral video is a supercut of John singing the song through the years since it was first released in 1970. While his performance is miraculously consistent, his wardrobe is anything but. (What exactly was that Donald Duck costume about?) So sit on the rooftop, kick off the moss, and enjoy.

And you can tell everybody...you saw it first.



Photo Credit Getty Images



Comments

View Comments



Terms of Use Privacy Statement Copyright and Trademark Notice EEO Public File Public Inspection File
iHeartRadio - Get the Free Radio App General Contest Guidelines
Advertise on News Radio 610 WTVN Get Results Contact Philanthropy
© 2017 iHeartMedia, Inc.

Listen Now on iHeartRADIO

Terms of Use Privacy Statement Copyright and Trademark Notice EEO Public File Public Inspection File
iHeartRadio - Get the Free Radio App General Contest Guidelines
Advertise on News Radio 610 WTVN Get Results Contact Philanthropy
© 2017 iHeartMedia, Inc.



Listen Now on iHeartRADIO

Web Page Segmentation

Listen Live on iHeartRADIO
On-Air News Podcasts & Media Connect Contests

News Radio 610 WTVN - News, Traffic, & Weather - Columbus, OH

Flashback: Supercut of Elton John singing 'Your Song' through the years

posted by Samantha Martin | Popdust - 4 years ago

comment share

Listen to Elton John on iHeartRadio

There's something exciting about being among the first few thousand people to watch a YouTube video on the verge of virality. Congrats! You're about to be one of those people (assuming you read this close to the publish date).

Elton John's "Your Song" is such a perfect marriage of singer to song. It's no wonder that it's one of the most popular wedding songs of all time. In a slew of major hits, "Your Song" manages to be John's signature song.

This soon-to-be viral video is a supercut of John singing the song through the years since it was first released in 1970. While his performance is miraculously consistent, his wardrobe is anything but. (What exactly was that Donald Duck costume about?) So sit on the rooftop, kick off the moss, and enjoy.

And you can tell everybody...you saw it first.



Photo Credit Getty Images



Comments

View Comments



Flashback: Supercut of Elton John singing 'Your Song' through the years

posted by Samantha Martin | Popdust - 4 years ago

comment share

Listen to Elton John on iHeartRadio

There's something exciting about being among the first few thousand people to watch a YouTube video on the verge of virality. Congrats! You're about to be one of those people (assuming you read this close to the publish date).

Elton John's "Your Song" is such a perfect marriage of singer to song. It's no wonder that it's one of the most popular wedding songs of all time. In a slew of major hits, "Your Song" manages to be John's signature song.

This soon-to-be viral video is a supercut of John singing the song through the years since it was first released in 1970. While his performance is miraculously consistent, his wardrobe is anything but. (What exactly was that Donald Duck costume about?) So sit on the rooftop, kick off the moss, and enjoy.

And you can tell everybody...you saw it first.



Photo Credit Getty Images



Comments

View Comments

Terms of Use Privacy Statement Copyright and Trademark Notice EEO Public File Public Inspection File

iHeartRadio - Get the Free Radio App General Contest Guidelines
Advertise on News Radio 610 WTVN Get Results Contact Philanthropy

© 2017 iHeartMedia, Inc.

Listen Now

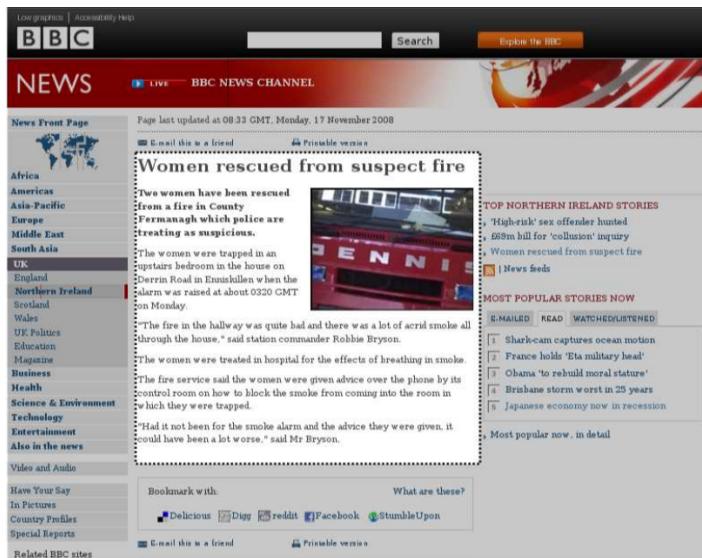
Terms of Use Privacy Statement Copyright and Trademark Notice EEO Public File Public Inspection File

iHeartRadio - Get the Free Radio App General Contest Guidelines
Advertise on News Radio 610 WTVN Get Results Contact Philanthropy

© 2017 iHeartMedia, Inc.

Listen Now

Web Page Segmentation: Downstream Tasks (Examples)



❑ Content Extraction

Image: *Language Independent Content Extraction from Web Pages*. Moreno et al., DIR'09.

Web Page Segmentation: Downstream Tasks (Examples)

The screenshot shows a news article titled "Women rescued from suspect fire". The article discusses two women being rescued from a fire in County Fermanagh. It includes a small image of a red Dennis fire truck, a sidebar with "TOP NORTHERN IRELAND STORIES", and a "MOST POPULAR STORIES NOW" section. The BBC navigation bar at the top includes links for "NEWS", "LIVE BBC NEWS CHANNEL", "Search", and "Explore the BBC".

The screenshot shows a music website for "LAUNCHcast plus". It features a search bar, navigation links for "SEARCH", "LAUNCHcast RADIO", "MUSIC VIDEOS", "ARTISTS", "STYLES", and "SONGS". A main headline reads "NOW! THE DIXIE CHICKS official artist club - get pre-sale concert tickets and more!". Below this are sections for "Sign Up For LAUNCHcast Plus", "LAUNCHcast plus", "Exclusive stations", "ARTIST SPOTLIGHT" featuring Norah Jones, and "MUSIC VIDEOS" with a thumbnail for "Dixie Chicks Artist Club". The page is branded with "YOUR YAHOO! MUSIC EXPERIENCE". A cyan box highlights the "Template Navigation Bar" on the right side of the page.

□ Content Extraction

Image: *Language Independent Content Extraction from Web Pages*. Moreno et al., DIR'09.

□ Template Detection

Image: *Automatic Data Extraction From Template Generated Web Pages*. Ma et al., PDPTA'03.

Template Navigation Table

Template Navigation Bar

Advertisement

Web Page Segmentation: Downstream Tasks (Examples)

The screenshot shows a news article titled "Women rescued from suspect fire". The article includes a photo of a red fire truck, a sidebar with "TOP NORTHERN IRELAND STORIES", and a "Most popular now" section. The BBC navigation bar at the top includes links for "Search", "Explore the BBC", and "BBC NEWS CHANNEL".

The screenshot shows a music website for "LAUNCHcast plus". It features sections for "Sign Up For LAUNCHcast Plus", "ARTIST SPOTLIGHT" (featuring Norah Jones), and "MUSIC VIDEOS" (featuring Dixie Chicks). The interface includes a search bar, navigation tabs for "SEARCH", "ALL", "GO", and links for "LAUNCHcast RADIO", "MUSIC VIDEOS", "ARTISTS", "STYLES", and "SONGS". A teal border highlights the "LAUNCHcast plus" header and the "ARTIST SPOTLIGHT" section.

Content Extraction

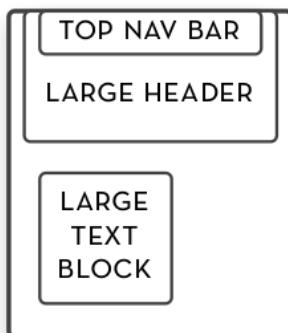
Image: *Language Independent Content Extraction from Web Pages*. Moreno et al., DIR'09.

Template Detection

Image: *Automatic Data Extraction From Template Generated Web Pages*. Ma et al., PDPTA'03.

Design Mining

Image: *Webzeitgeist: Design Mining the Web*. Kumar et al., CHI'13.



LAYOUT QUERY

The screenshot shows a website for "Ayutthaya" featuring a large image of a Buddha statue. The site includes sections for "WEBSITE", "WEB DESIGN", and "WEB MARKETING".

The screenshot shows a website for "IMITATION" featuring a cartoon of sheep. The site includes sections for "DESIGN & BUILD" and "THINK BIG".

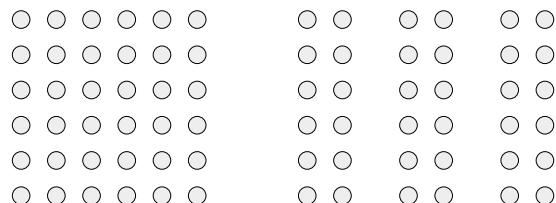
Template Navigation Table
Template Navigation Bar
Advertisement

Concept Formation: Web Page Segment

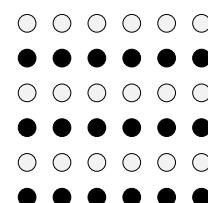
A web page segment is a part of a web page containing those elements that belong together as per agreement among a majority of viewers.

Rationale: Web pages are created for human viewers, and so are segments

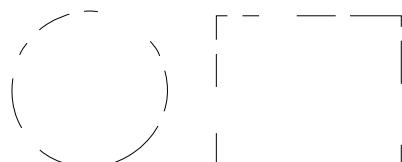
Gestalt Principles provide common ground



Proximity



Similarity



Closure



Symmetry

Evaluation Framework for Web Page Segmentation

A web page segment is a part of a web page containing those elements that belong together as per agreement among a majority of viewers.

Elements $E = \{e_1, \dots, e_n\}$

Segmentation $S = \{s_1, \dots, s_m\}$ with segments $s_i \subseteq E$

Evaluation Framework for Web Page Segmentation

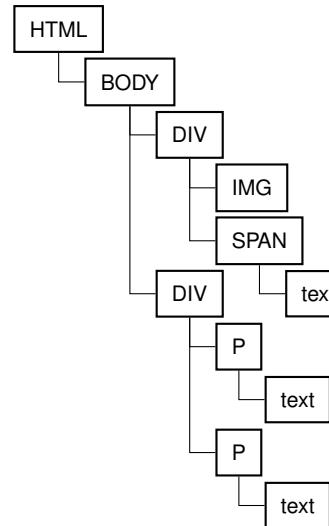
A web page segment is a part of a web page containing those elements that belong together as per agreement among a majority of viewers.

Elements $E = \{e_1, \dots, e_n\}$

Segmentation $S = \{s_1, \dots, s_m\}$ with segments $s_i \subseteq E$

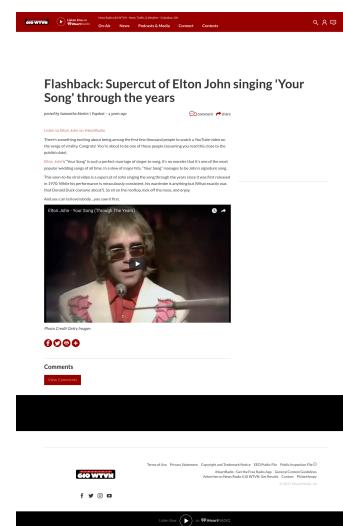
Suggested sets of elements:

Listen Live on iHea
rtRADIO News Radi
o 610WTVN-News, T
raffic, Weather - C
olumbus, OH On-Air
News Podcasts Me
dia Connect Contes
ts Flashback: Supe
rcur of Elton John
singing 'Your Song'
through the years
posted by Samantha
Martin | Popdust -
4 years ago comme
nt share Listen to
Elton John on iHear
tRadio There's som
ething exciting abo
ut being among the

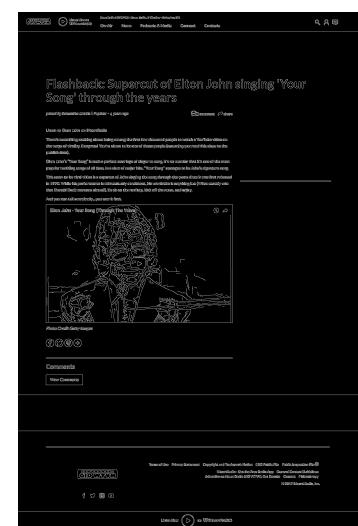


Characters

DOM nodes



Pixels



Edges

Evaluation Framework for Web Page Segmentation

A web page segment is a part of a web page containing those elements that belong together as per agreement among a majority of viewers.

Elements $E = \{e_1, \dots, e_n\}$

Segmentation $S = \{s_1, \dots, s_m\}$ with segments $s_i \subseteq E$

Precision

$$P_{B^3}(S, S^*) = \text{avg}_e \left(\frac{|\text{elements in same segment as } e \text{ in both } S \text{ and } S^*|}{|\text{elements in same segment as } e \text{ in } S|} \right)$$

Recall

$$R_{B^3}(S, S^*) = \text{avg}_e \left(\frac{|\text{elements in same segment as } e \text{ in both } S \text{ and } S^*|}{|\text{elements in same segment as } e \text{ in } S^*|} \right)$$

F-Measure, F_{B^3} , is defined as the harmonic mean of precision and recall as usual

Note: $P_{B^3}(S, S') = R_{B^3}(S', S) \Rightarrow F_{B^3}(S, S') = F_{B^3}(S', S)$

Evaluation Framework for Web Page Segmentation

posted by Sonnetto_Melts | Popular - 4 years ago

Listen to Elton John on [HeartRadio](#)

There's something exciting about being among the first few thousand people to watch a YouTube video on the most unlikely day. Congrats! You're about to be one of those people, deserving of such close to the published date.

Elton John's "Your Song" is such a perfect marriage of singer to song. It's no wonder that it's one of the most popular wedding songs of all time. In a slew of major hits, "Your Song" manages to be John's signature song. This is the same John who was the first to play piano on stage at the Royal Albert Hall in 1970, and the same John who recorded his first album in 1970. While his performance is miraculously consistent, his wardrobe is anything but. What exactly was that Donald Duck costume about? So sit on the rooftop, kick off the roses, and enjoy.

And you can tell everybody... you saw it first.

Photo Credit: Getty Images

[Comment](#) [Share](#)

[Comments](#) [View Comments](#)

[Listen Now](#) [on iHeartRadio](#)

Terms of Use Privacy Statement Copyright and Trademark Notice EEO Public File Public Inspection File D
iHeartRadio - Get the Free Radio App General Content Guidelines Advertisers Terms of Service © 2017 iHeartMedia, Inc.



posted by Sonnetto_Melts | Popular - 4 years ago

Listen to Elton John on [HeartRadio](#)

There's something exciting about being among the first few thousand people to watch a YouTube video on the most unlikely day. Congrats! You're about to be one of those people, deserving of such close to the published date.

Elton John's "Your Song" is such a perfect marriage of singer to song. It's no wonder that it's one of the most popular wedding songs of all time. In a slew of major hits, "Your Song" manages to be John's signature song. This is the same John who was the first to play piano on stage at the Royal Albert Hall in 1970, and the same John who recorded his first album in 1970. While his performance is miraculously consistent, his wardrobe is anything but. What exactly was that Donald Duck costume about? So sit on the rooftop, kick off the roses, and enjoy.

And you can tell everybody... you saw it first.

Photo Credit: Getty Images

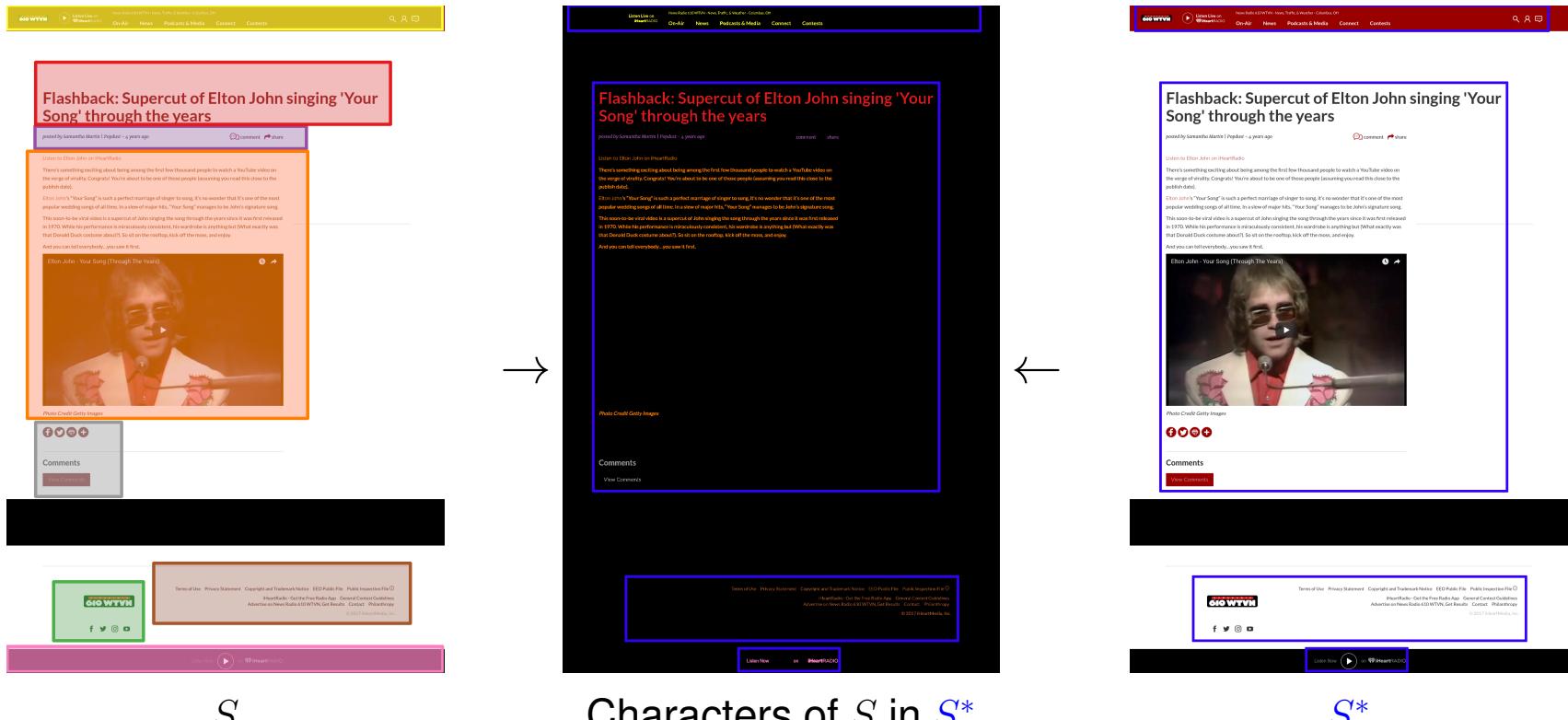
[Comment](#) [Share](#)

[Comments](#) [View Comments](#)

[Listen Now](#) [on iHeartRadio](#)

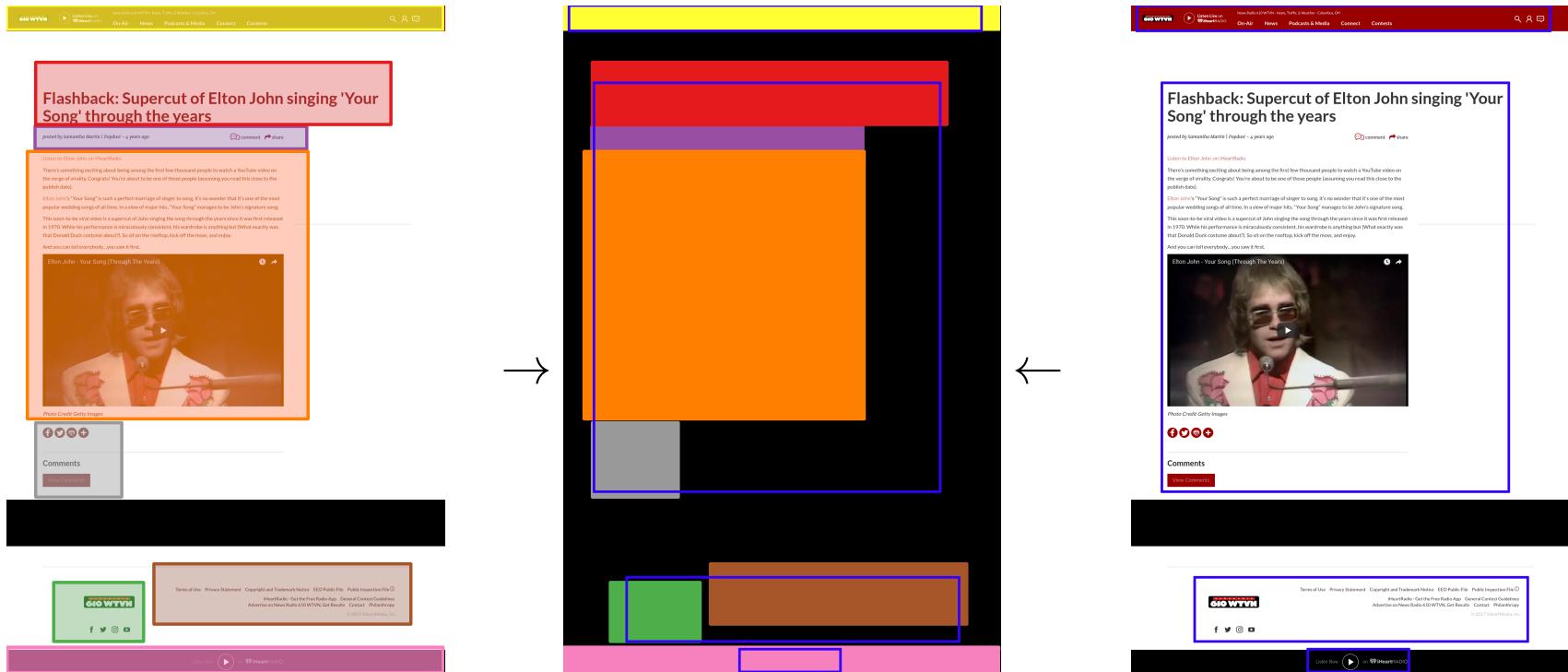
Terms of Use Privacy Statement Copyright and Trademark Notice EEO Public File Public Inspection File D
iHeartRadio - Get the Free Radio App General Content Guidelines Advertisers Terms of Service © 2017 iHeartMedia, Inc.

Evaluation Framework for Web Page Segmentation



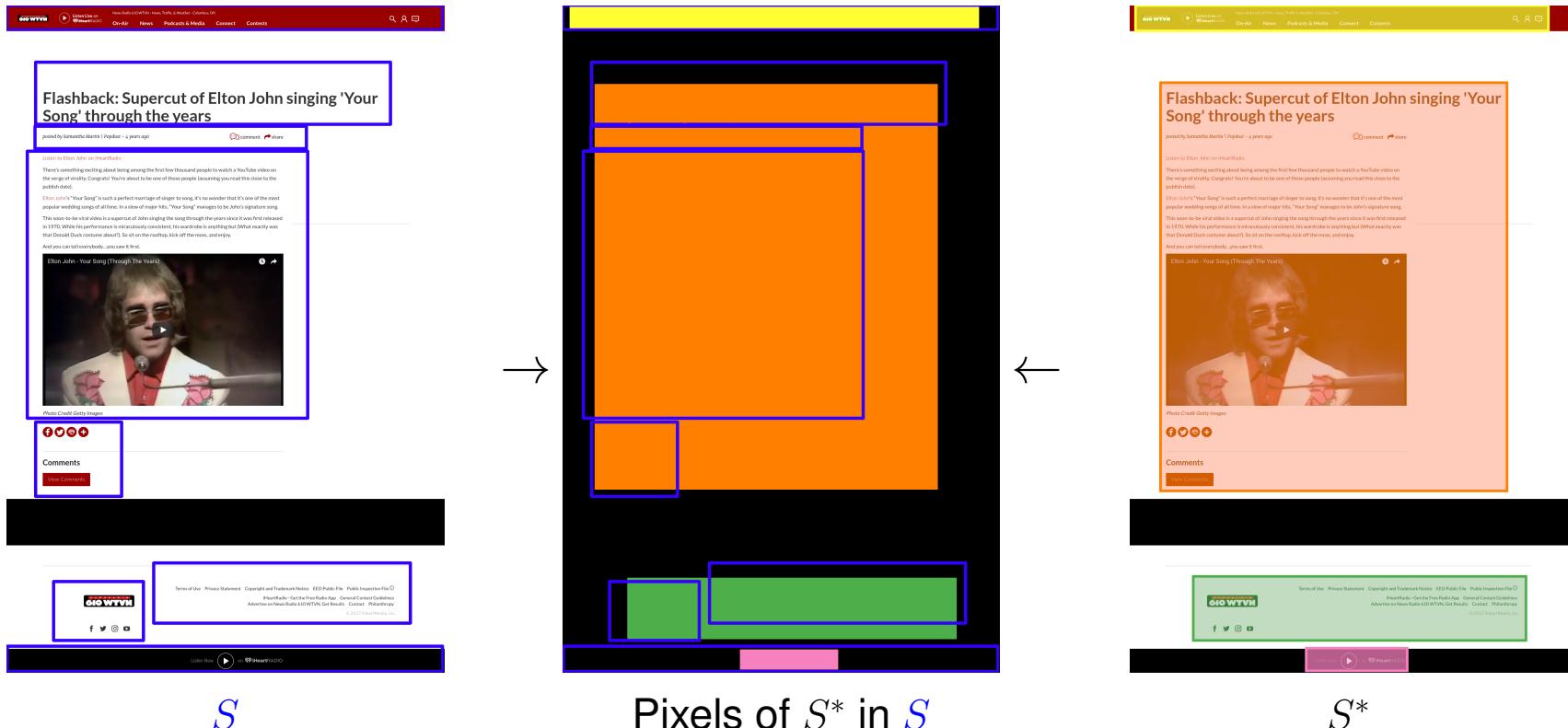
$$P_{B^3}(S, S^*) = \text{avg}_e \left(\frac{|\text{elements in same segment as } e \text{ in both } S \text{ and } S^*|}{|\text{elements in same segment as } e \text{ in } S|} \right)$$

Evaluation Framework for Web Page Segmentation



$$P_{B^3}(S, S^*) = \text{avg}_e \left(\frac{|\text{elements in same segment as } e \text{ in both } S \text{ and } S^*|}{|\text{elements in same segment as } e \text{ in } S|} \right)$$

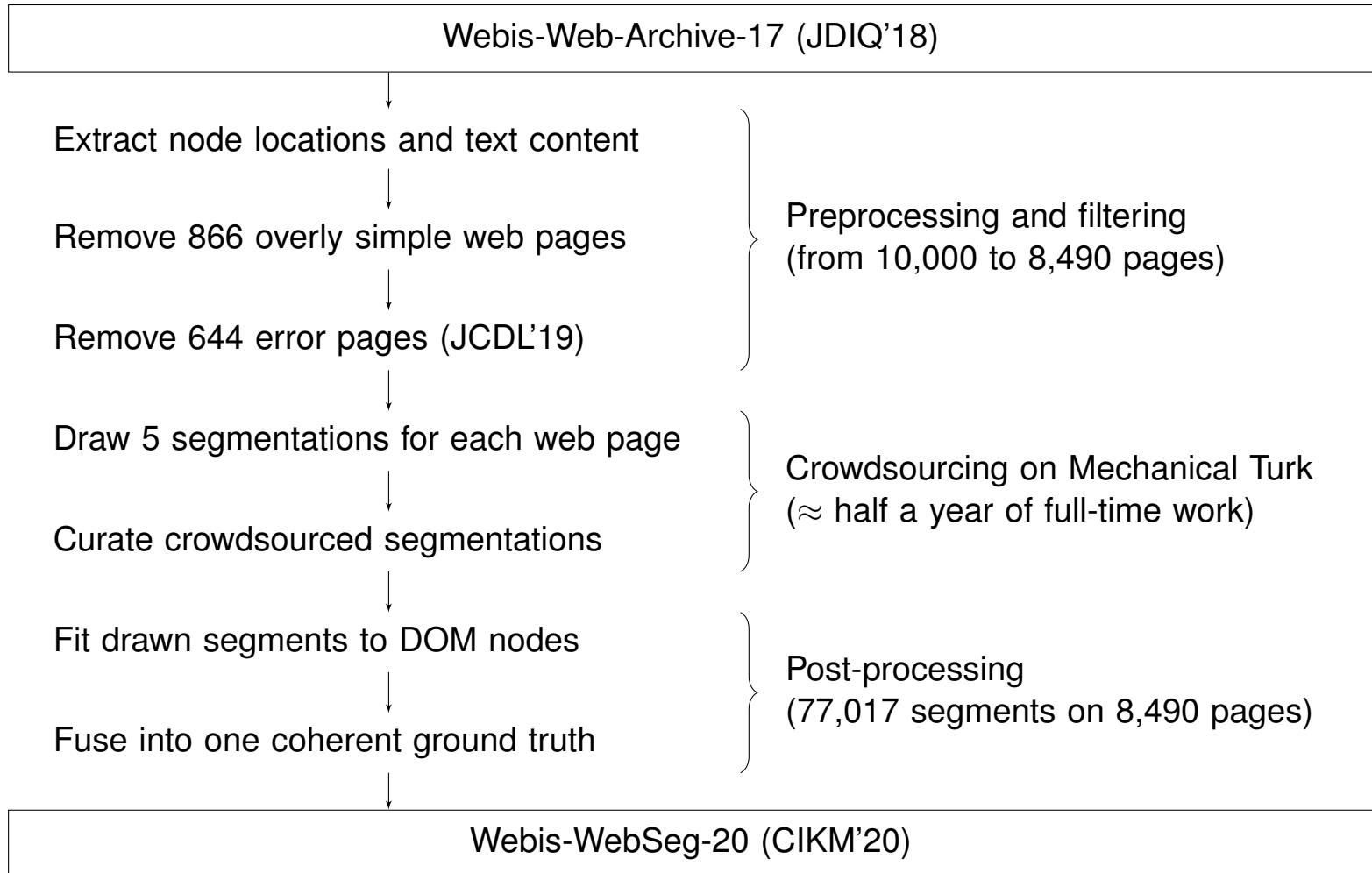
Evaluation Framework for Web Page Segmentation



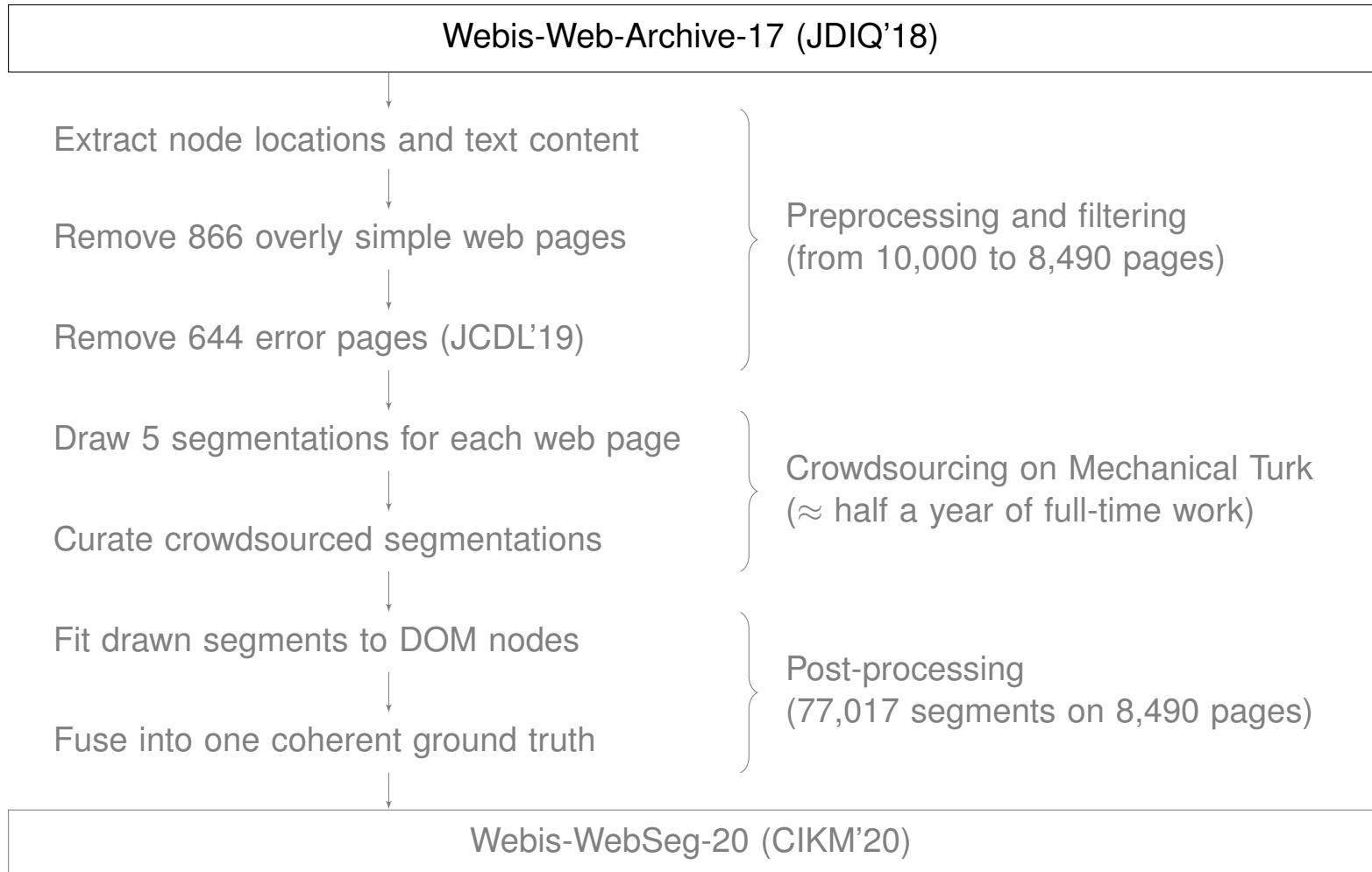
$$P_{B^3}(S, S^*) = \text{avg}_e \left(\frac{|\text{elements in same segment as } e \text{ in both } S \text{ and } S^*|}{|\text{elements in same segment as } e \text{ in } S|} \right)$$

$$R_{B^3}(S, S^*) = \text{avg}_e \left(\frac{|\text{elements in same segment as } e \text{ in both } S \text{ and } S^*|}{|\text{elements in same segment as } e \text{ in } S^*|} \right)$$

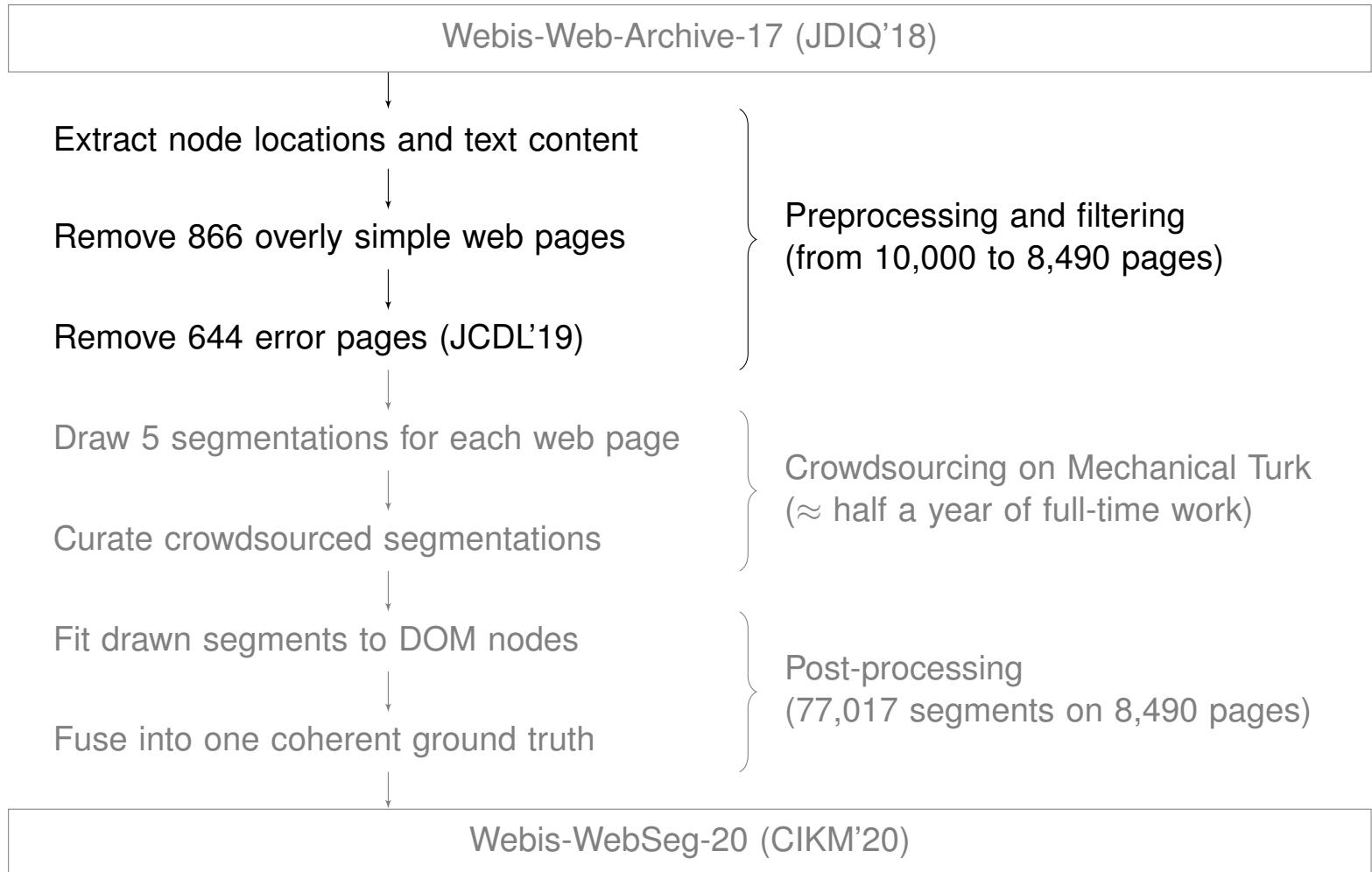
The Webis-WebSeg-20 Dataset



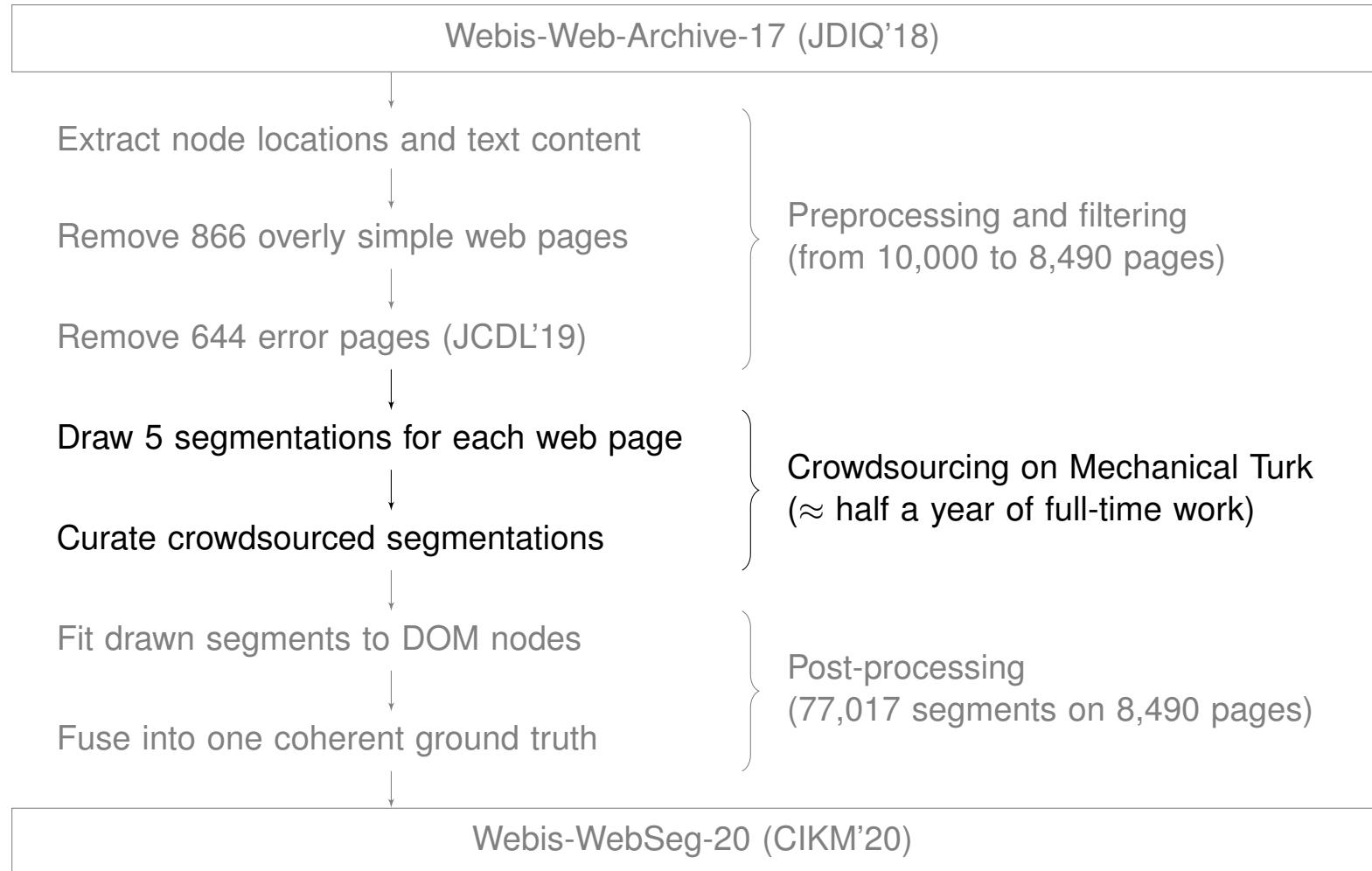
The Webis-WebSeg-20 Dataset



The Webis-WebSeg-20 Dataset



The Webis-WebSeg-20 Dataset



The Webis-WebSeg-20 Dataset

Webis-Web-Archive-17 (JDIQ'18)

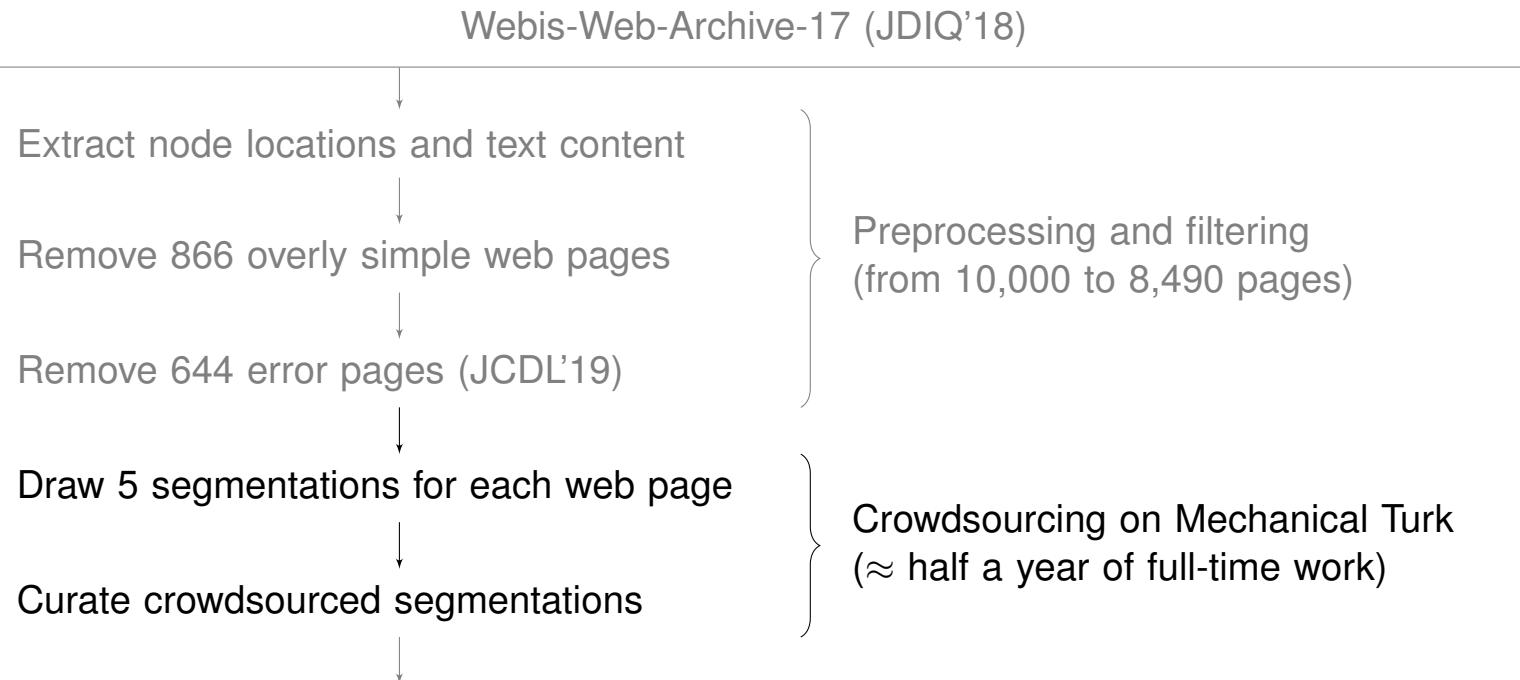
Extract node locations and text content

The screenshot shows a web page from the Vietnamese news site VietBao. At the top, there's a search bar, a date/time indicator (Thứ sáu, 22/09/2017 | 05:40 GMT+7), and a weather widget (Hà Nội, 31°C). A sidebar on the right displays exchange rates (Giá Vàng, Tỷ Giá, Xem Ngày 1/08 Âm Lịch) and a calendar (Ngày, tháng, Năm). The main navigation menu includes links like Việt Báo, Chính Trị, Thế Giới, Xã Hội, Kinh Doanh, Pháp Luật, Bóng Đá, Giải Trí, Sống Trẻ, Đời Sống, Sức Khỏe, and three more hidden under a dropdown.

The central content area features a banner for "Thế Giới Giải Trí Chàng Nặng" with three images: a woman in a red dress, a woman in a white dress, and a couple. Below this is a large advertisement for a lottery sweepstakes (SUPER! Das ist kein Scherz! Sie sind unser 1.000.000ster Besucher! Unser Zufallsystem der möglichen Gewinner könnte Sie als möglichen Gewinner von FANTASTISCHEN APPLE Produkten ziehen.) with a red circle highlighting the number 1.000.000ster. Another ad for "Klicken Sie hier" is visible on the right.

The main article headline is "Bạn gái cặp với người nước ngoài vì tiền" (Hot SỰ KIỆN NÓNG). It includes social sharing icons (Facebook, Twitter, etc.), a timestamp (Thứ năm, 27 Tháng Ba 2009, 09:36 GMT+7), and a video player with a "Subscribe" button. The article text discusses a relationship between a Vietnamese girl and an overseas man for money. To the right, there's a sidebar titled "ĐỌC NHIỀU" featuring other news items like "Đam Vĩnh Hưng bất ngờ lên tiếng vụ việc của danh hài Xuân Bắc".

The Webis-WebSeg-20 Dataset



Agreement measure

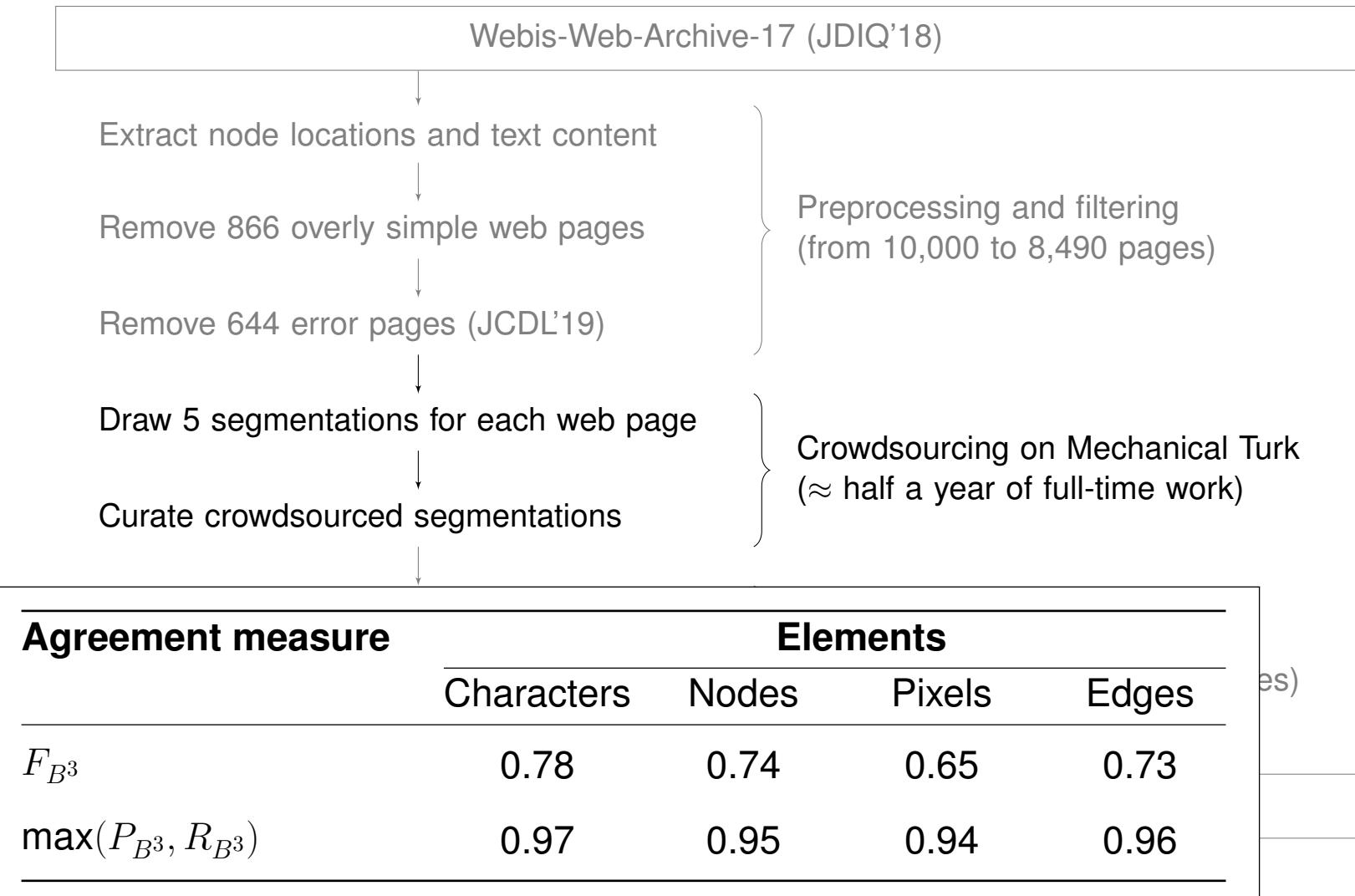
Elements

	Characters	Nodes	Pixels	Edges
--	------------	-------	--------	-------

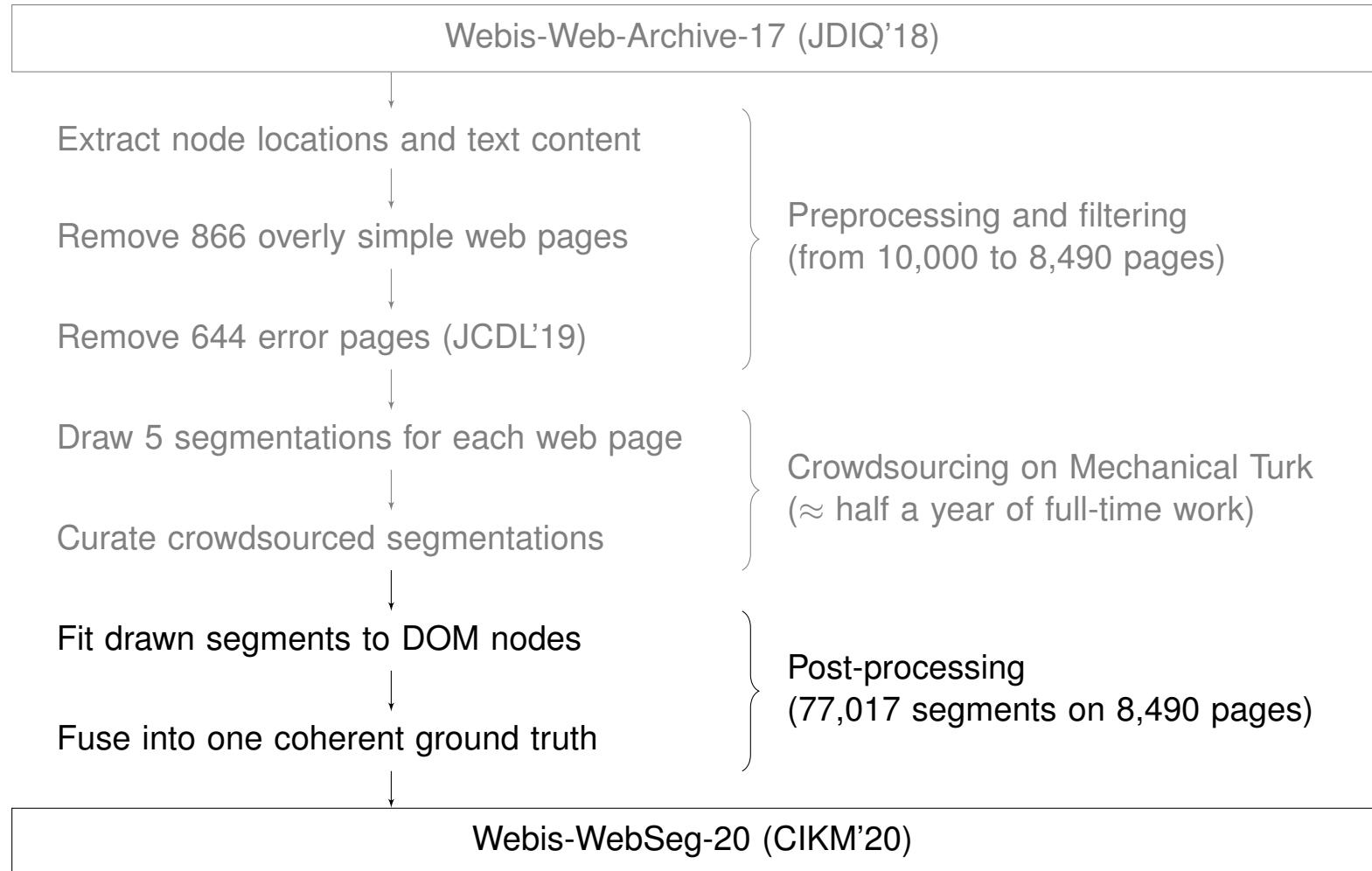
F_{B^3}	0.78	0.74	0.65	0.73
-----------	------	------	------	------

es)

The Webis-WebSeg-20 Dataset



The Webis-WebSeg-20 Dataset



Web Page Segmentation Revisited: Evaluation Framework and Dataset

Evaluation Framework for Web Page Segmentation

- Segmentation similarity, quality, and fusion
- Comparison of level of detail of segmentations
- Adjustable for different downstream tasks

Webis-WebSeg-20

- 8,490 pages from 4,824 sites
- 5 human annotators each page
- Segments in “Simple Feature Access” standard
- Web pages provided in several representations:
 - HTML file
 - Screenshot
 - Screenshot coordinates of DOM nodes
 - Webis-Web-Archive-17 WARC file



https://webis.de/publications.html?q=johannes+kiesel+web+archive#stein_2020w

Paper, browser, code, data