

An Empirical Comparison of Web Page Segmentation Algorithms

ECIR 2021



Johannes
Kiesel¹



Lars
Meyer¹



Florian
Kneist¹



Benno
Stein¹



Martin
Potthast²

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.

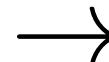


Photo Credit Getty Images



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

WTVP News Radio 610 WTVP - News, Traffic, & Weather - Columbus, OH
Listen Live on iHeartRadio
On-Air News Podcasts & Media Connect Contests
Search

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

WTVP News Radio 610 WTVP - News, Traffic, & Weather - Columbus, OH
Listen Live on iHeartRadio
On-Air News Podcasts & Media Connect Contests
Search

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 WTVP, 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 WTVP, Get Results Contact Philanthropy
© 2017 iHeartMedia, Inc.

Listen Now on iHeartRadio

Web Page Segmentation

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 | Podpost - 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 | Podpost - 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

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

Search icon, magnifying glass icon, and a small message 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



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

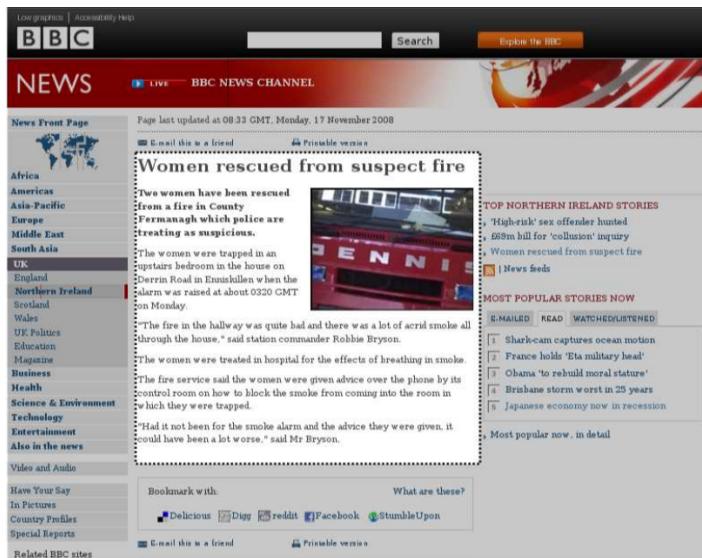
Terms of Use Privacy Statement Copyright and Trademark Notice EEO Public File Public Inspection File

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: Downstream Tasks (Examples)



Content Extraction

Image: *Language Independent Content Extraction from Web Pages*. Javier 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 fire truck and several paragraphs of text. The BBC navigation bar is visible at the top, along with a sidebar containing links to various news categories like Africa, Americas, Asia-Pacific, Europe, Middle East, South Asia, UK, and more.

The screenshot shows the LAUNCHcast plus website, which is part of the Yahoo! Music Experience. It features a search bar, navigation links for SEARCH, ALL, MUSIC VIDEOS, ARTISTS, STYLES, and SONGS, and a sign-up form for LAUNCHcast Plus. The main content area highlights "LAUNCHcast plus" and lists "Exclusive stations" such as 1980's Hair Flare, Martini Lounge, Affectionate, One Hit Wonders, Show Tunes, Classic FM, and Women of 1995. There are also sections for "ARTIST SPOTLIGHT" featuring Norah Jones and "MUSIC VIDEOS" showing clips for Dixie Chicks and Britney Spears.

Template Navigation Table
Template Navigation Bar
Advertisement

□ Content Extraction

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

□ Template Detection

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

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", "LAUNCHcast FREE", "LAUNCHcast plus", "ARTIST SPOTLIGHT", and "MUSIC VIDEOS". The "LAUNCHcast plus" section highlights "Exclusive stations" like "1980's Hair Flare" and "Hip-Hop". The "ARTIST SPOTLIGHT" section features Norah Jones. The "MUSIC VIDEOS" section shows a video thumbnail for "Dixie Chicks Artist Club". The page includes a search bar and navigation links for "SEARCH", "LAUNCHcast RADIO", "MUSIC VIDEOS", "ARTISTS", "STYLES", and "SONGS". A teal border surrounds the right side of the page, containing labels for "Template Navigation Bar", "Advertisement", and "Template Navigation Form".

□ Content Extraction

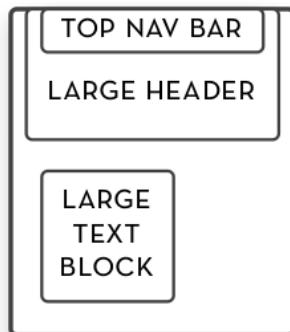
Image: *Language Independent Content Extraction from Web Pages*. Javier et al., DIR'09.

□ 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". It features a large image of a Buddha statue, a "WE CREATE GREAT VISUAL STORIES THAT BRINGS PEOPLE TO YOUR SITE" call-to-action, and sections for "WEB DESIGN", "UI/UX", and "DESIGN & BUILD".

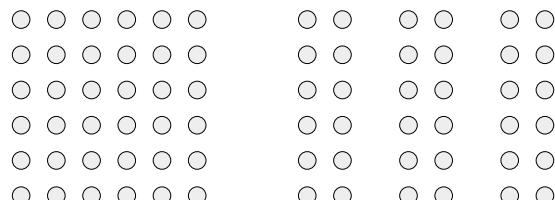
The screenshot shows a website for "IMITATION". It features a cartoon illustration of sheep, a "DESIGN & BUILD" section, and a "WEBSITE DESIGN" section. The page includes a "GIVE CREDIT", "BE COOL", "ASK FIRST", and "THINK BIG" footer.

Concept Formation: Web Page Segment

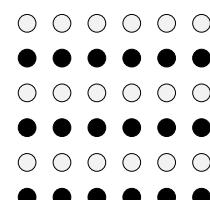
A web page segment is a part of a web page containing those elements that belong together as per agreement among a majority of viewers.

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

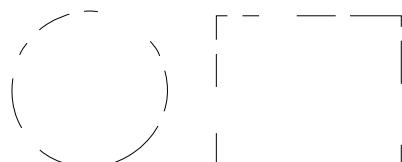
Gestalt Principles provide common ground



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

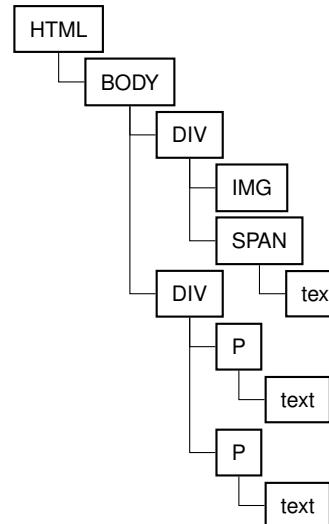
A web page segment is a part of a web page containing those elements that belong together as per agreement among a majority of viewers.

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

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

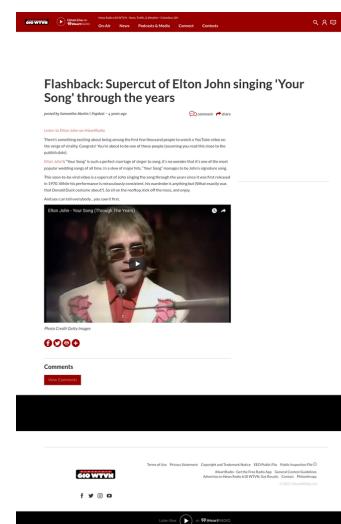
Suggested sets of elements:

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

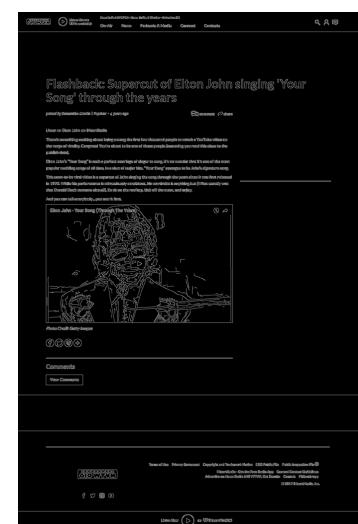


Characters

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 kieseljohannes | Popular - 4 years ago

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 most popular day. Congrats! You're about to be one of those people. (assuming you read this close to the publication 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 first time I've seen a full-length supercut of the song, and I'm so happy that it was released in 1970. While his performance is miraculously consistent, his wardrobe is anything but. Elton nearly was that Donald Duck costume about 5s in at the beginning, kick off the mous, and enjoy.

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

Elton John - Your Song (Through The Years)

Photo Credit Getty Images

Comments
View Comments

GIG WTVN
f @iheartradio

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



posted by kieseljohannes | Popular - 4 years ago

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 most popular day. Congrats! You're about to be one of those people. (assuming you read this close to the publication 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 first time I've seen a full-length supercut of the song, and I'm so happy that it was released in 1970. While his performance is miraculously consistent, his wardrobe is anything but. Elton nearly was that Donald Duck costume about 5s in at the beginning, kick off the mous, and enjoy.

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

Elton John - Your Song (Through The Years)

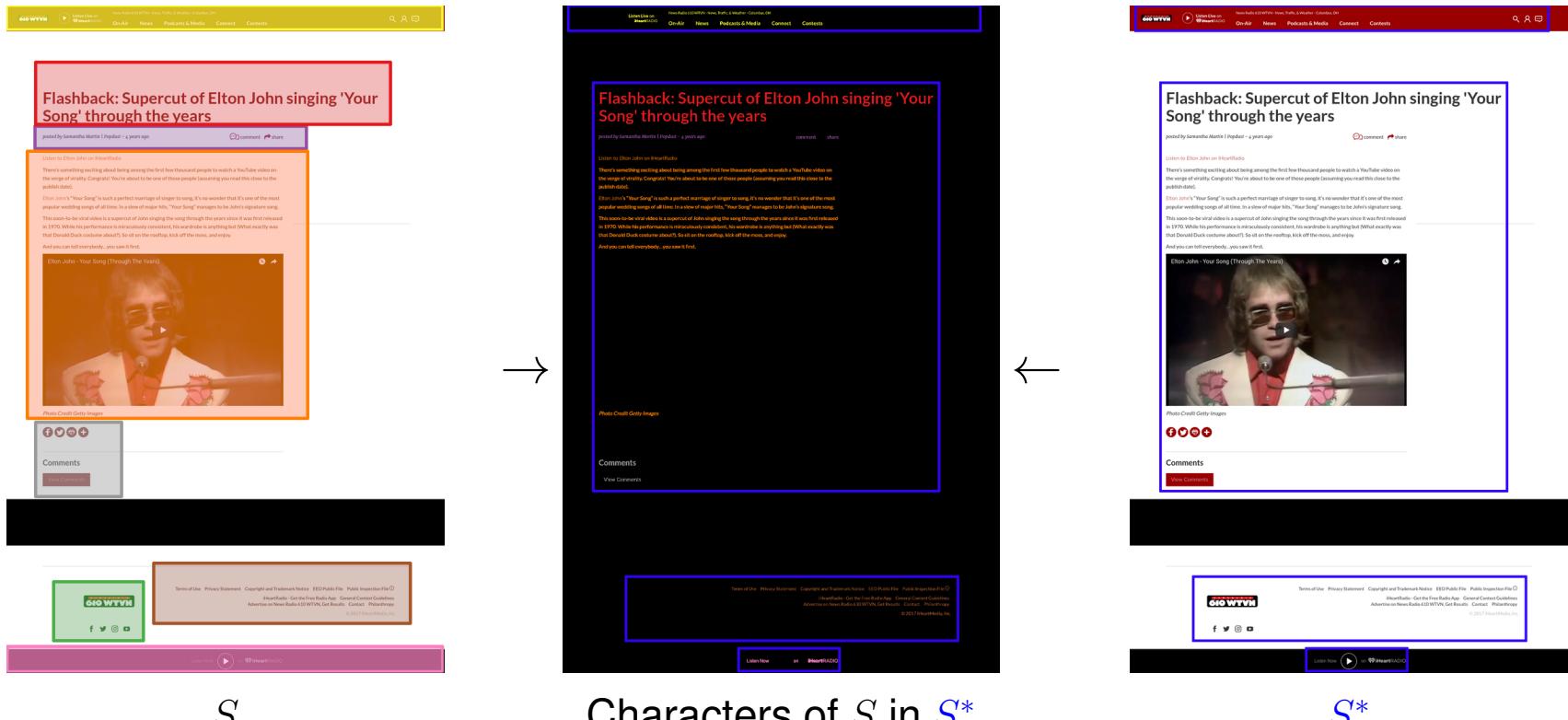
Photo Credit Getty Images

Comments
View Comments

GIG WTVN
f @iheartradio

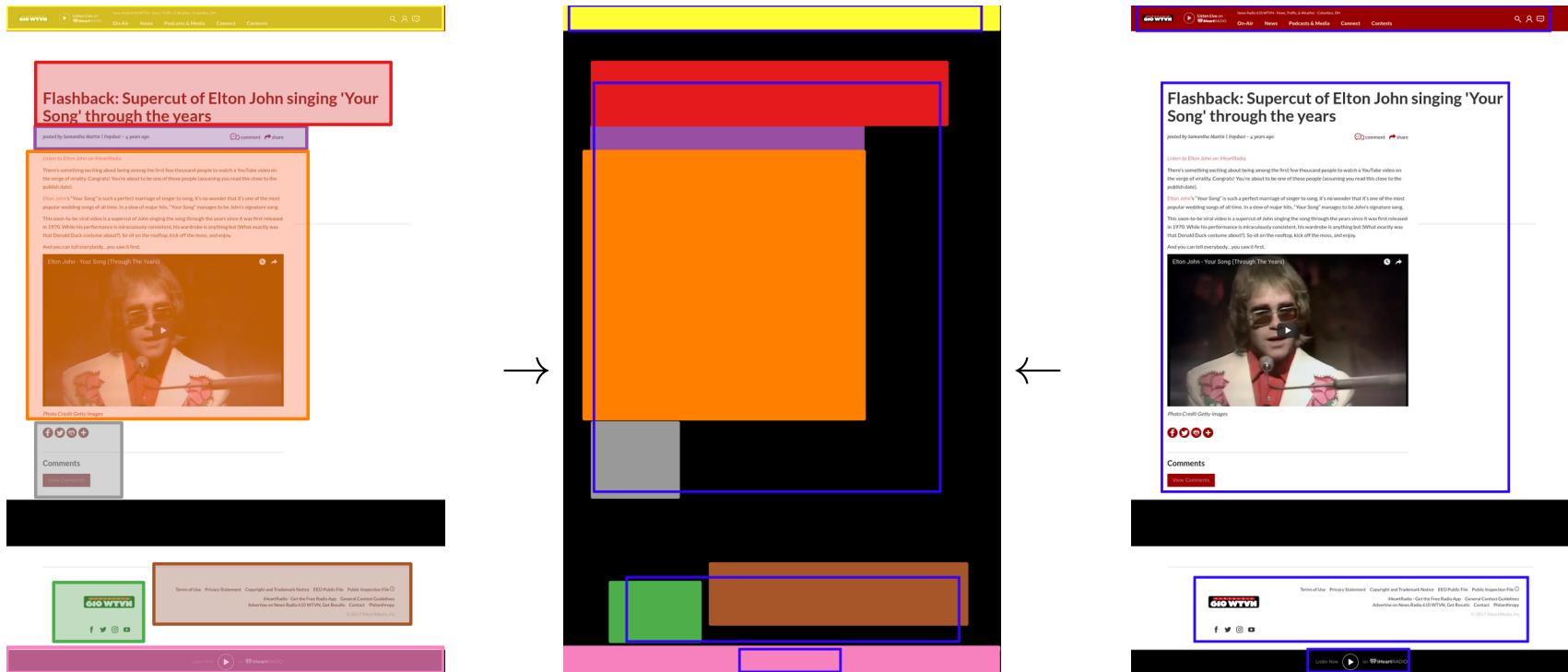
Terms of Use Privacy Statement Copyright and Trademark Notice EEO Public File Public Inspection File iHeartRadio - Get the Free Radio App General Content Guidelines Advertiser Information © 2017 iHeartMedia, Inc.

Evaluation Framework for Web Page Segmentation



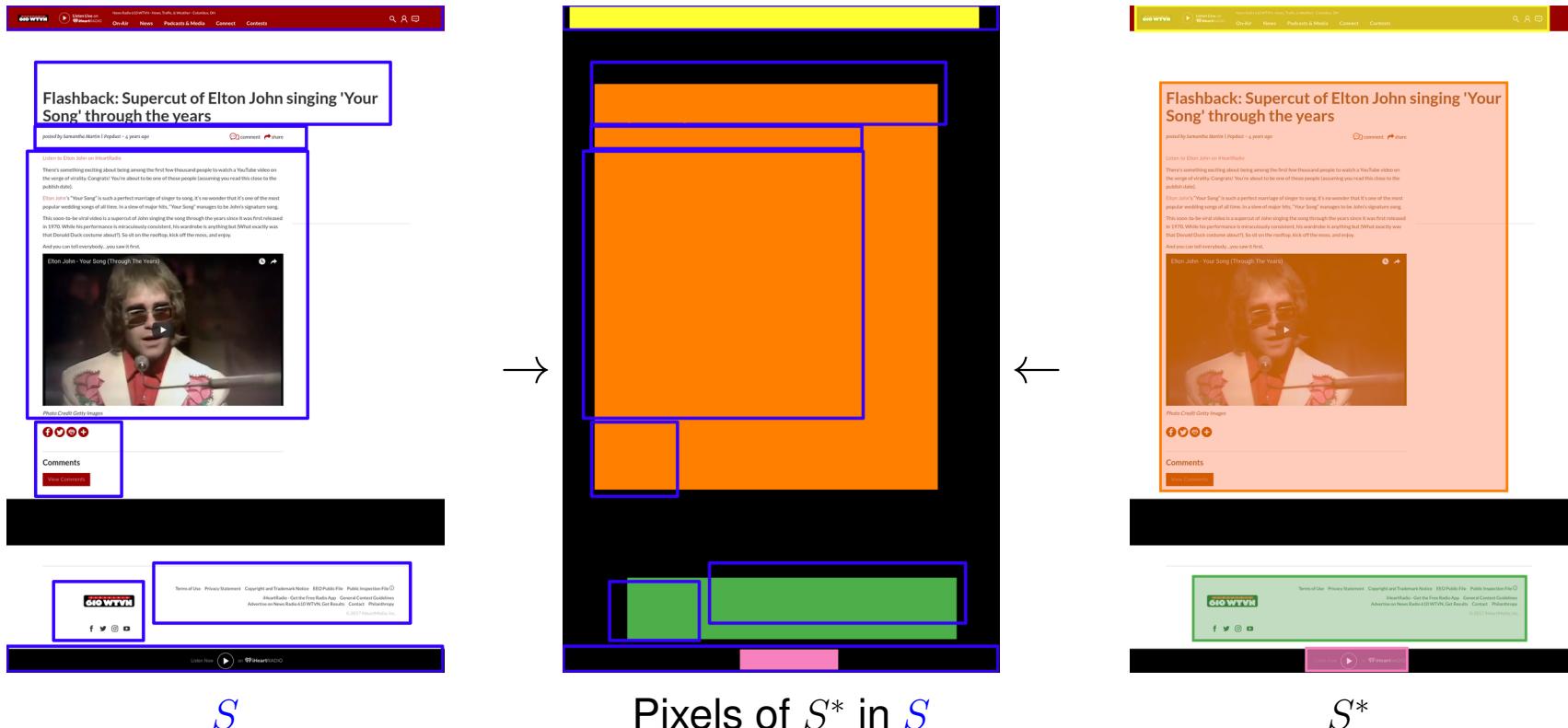
$$P_{B^3}(S, S^*) = \text{avg}_e \left(\frac{|\text{elements in same segment as } e \text{ in both } S \text{ and } S^*|}{|\text{elements in same segment as } e \text{ in } S|} \right)$$

Evaluation Framework for Web Page Segmentation



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

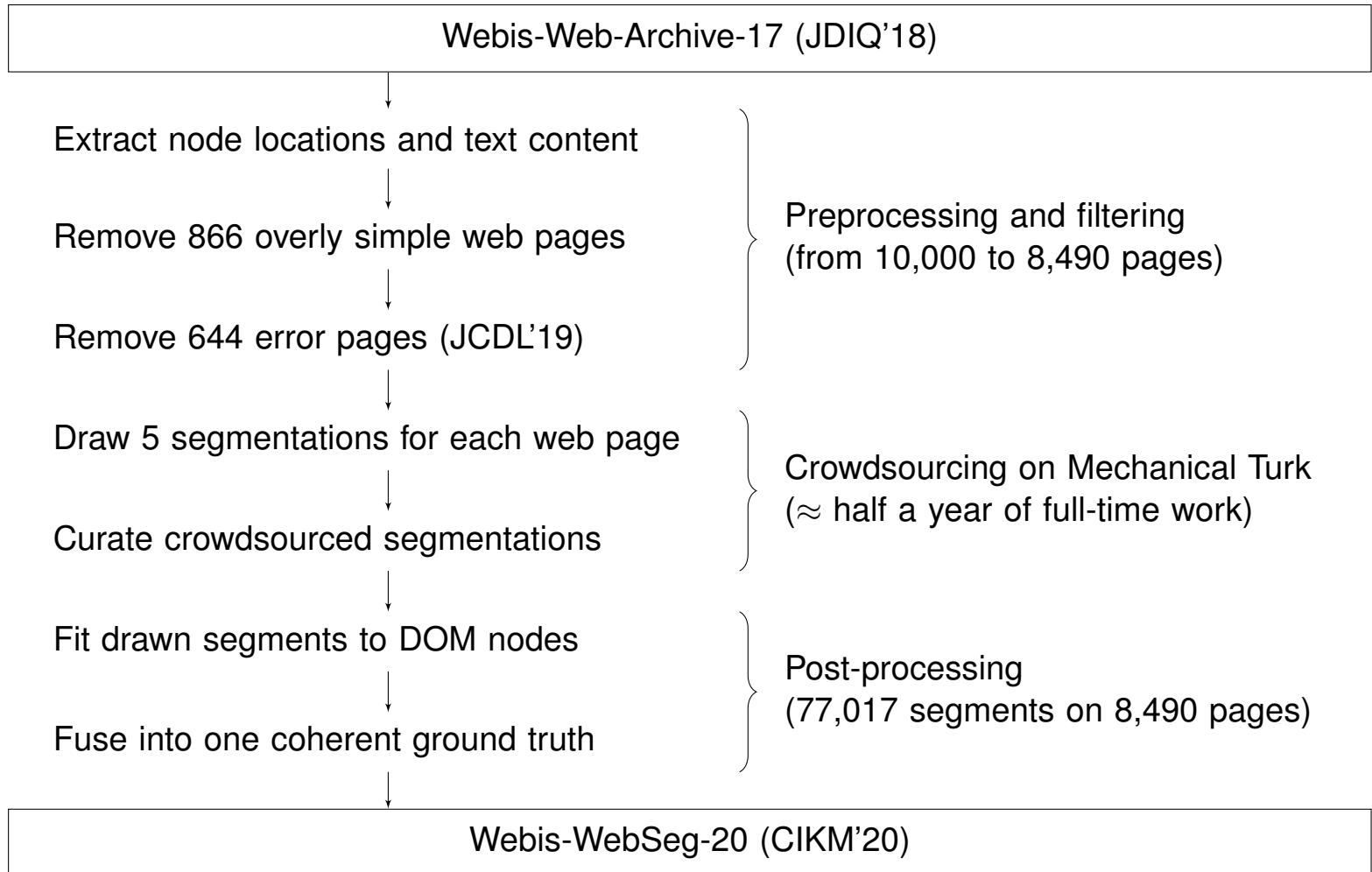
Evaluation Framework for Web Page Segmentation



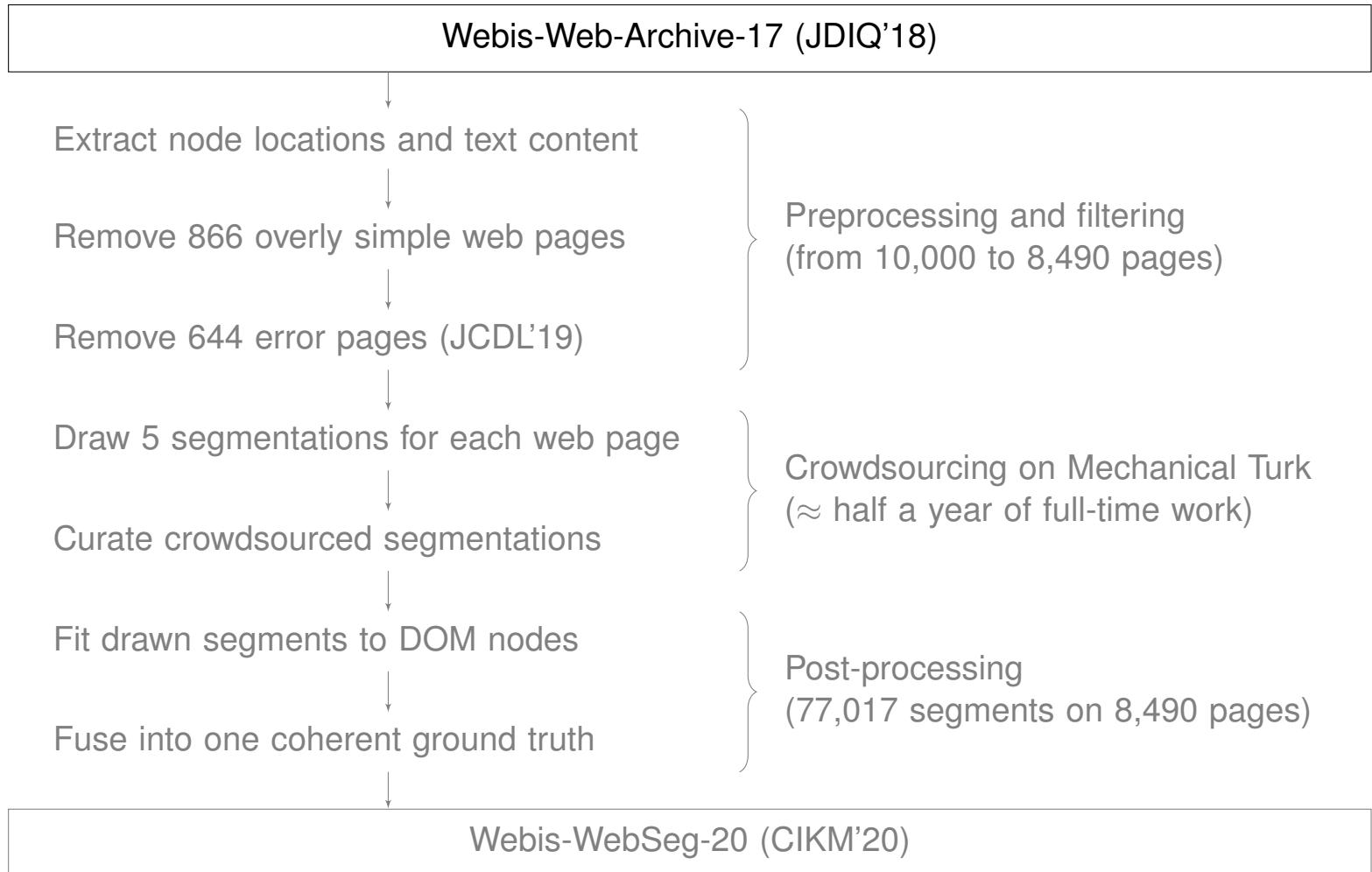
$$P_{B^3}(S, S^*) = \text{avg}_e \left(\frac{|\text{elements in same segment as } e \text{ in both } S \text{ and } S^*|}{|\text{elements in same segment as } e \text{ in } S|} \right)$$

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

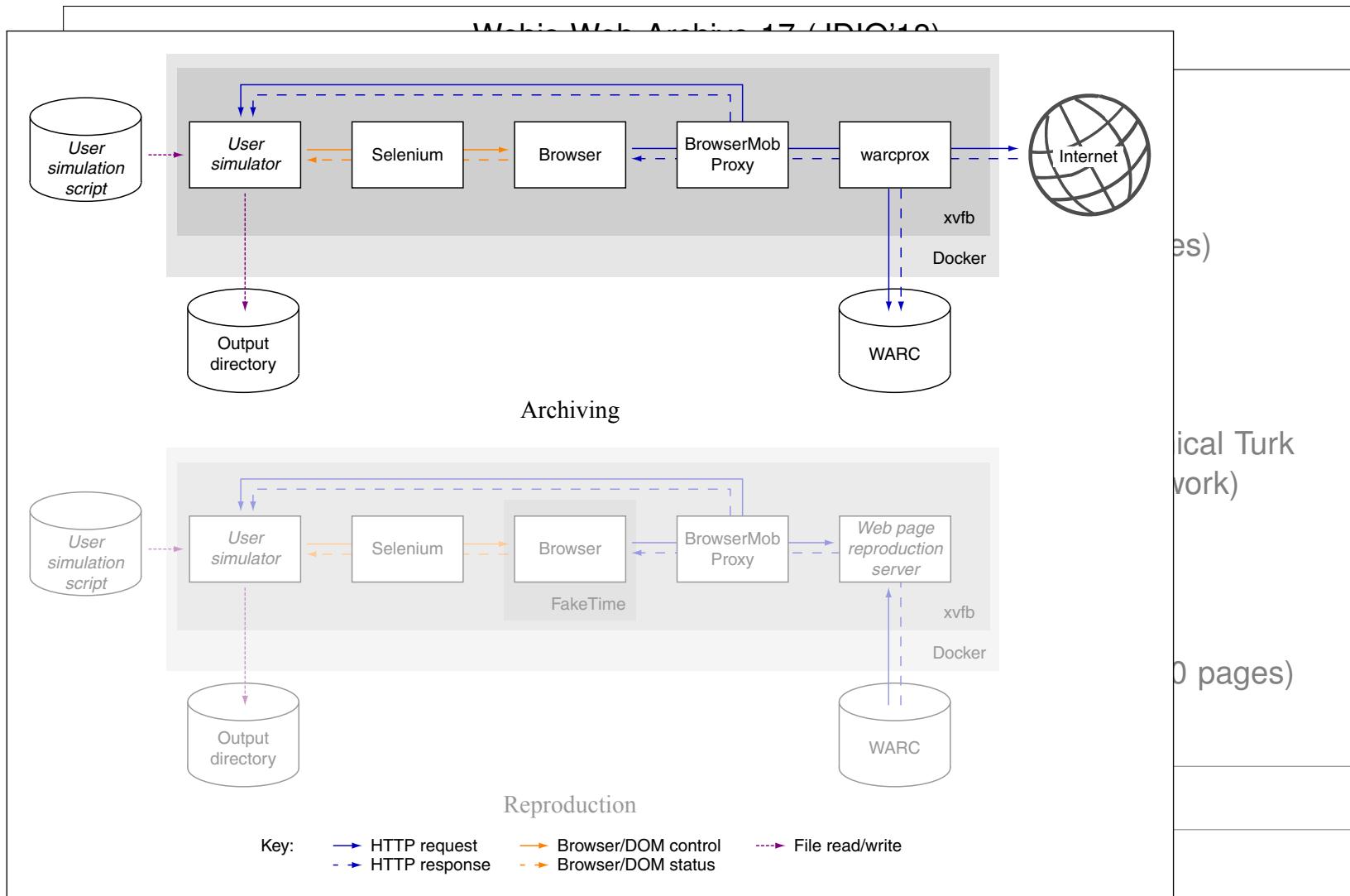
The Webis-WebSeg-20 Dataset



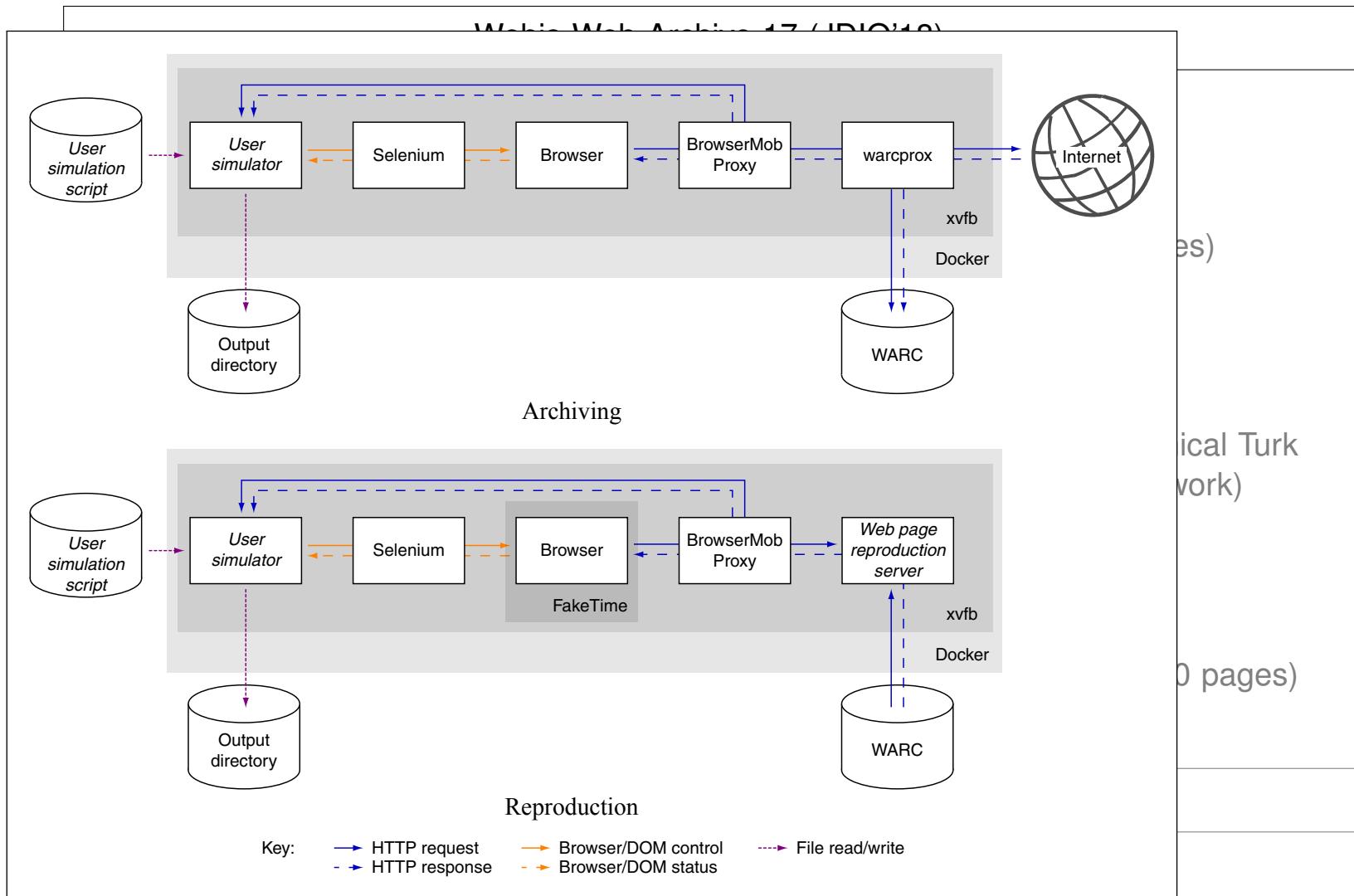
The Webis-WebSeg-20 Dataset



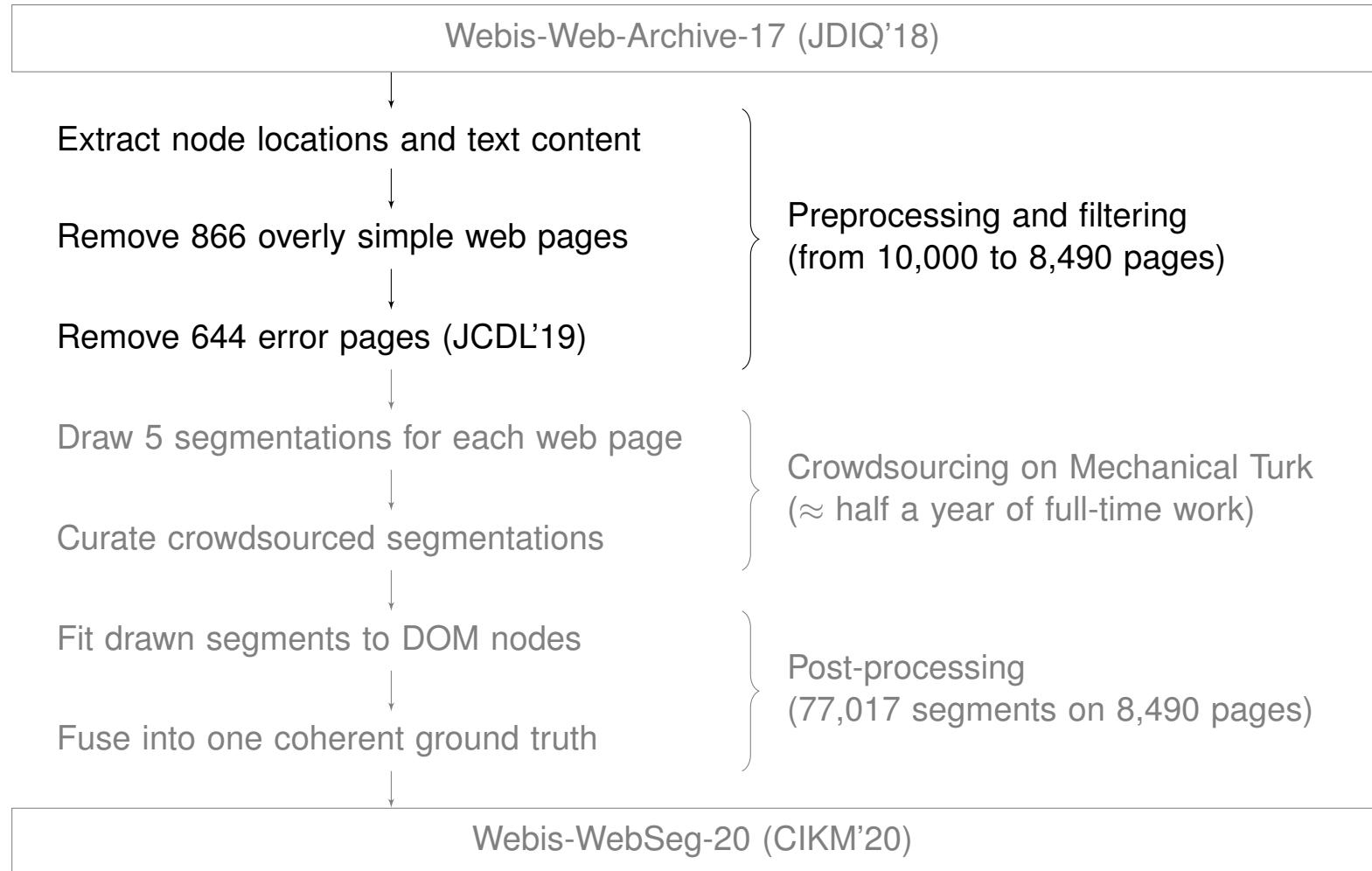
The Webis-WebSeg-20 Dataset



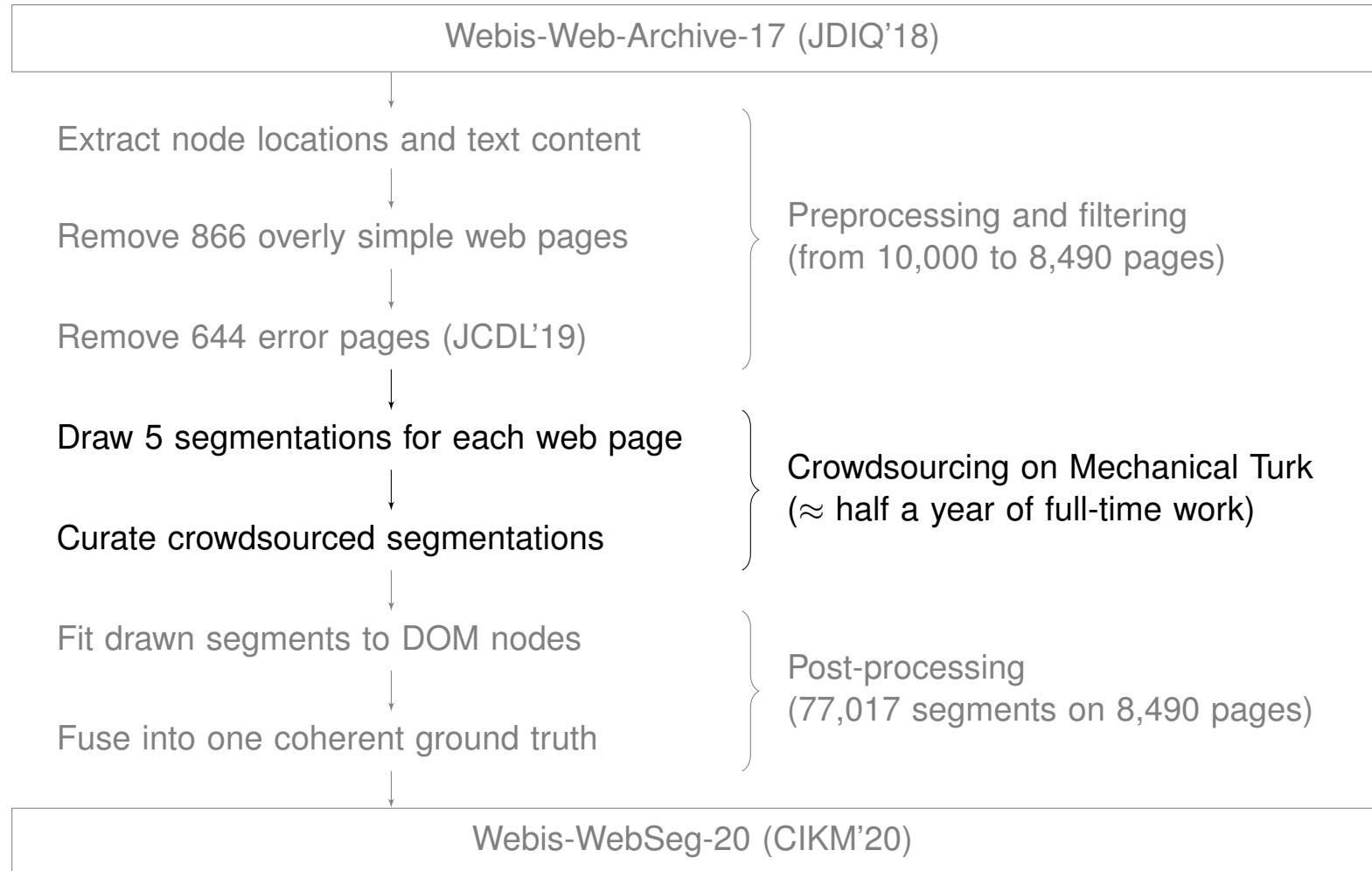
The Webis-WebSeg-20 Dataset



The Webis-WebSeg-20 Dataset



The Webis-WebSeg-20 Dataset



The Webis-WebSeg-20 Dataset

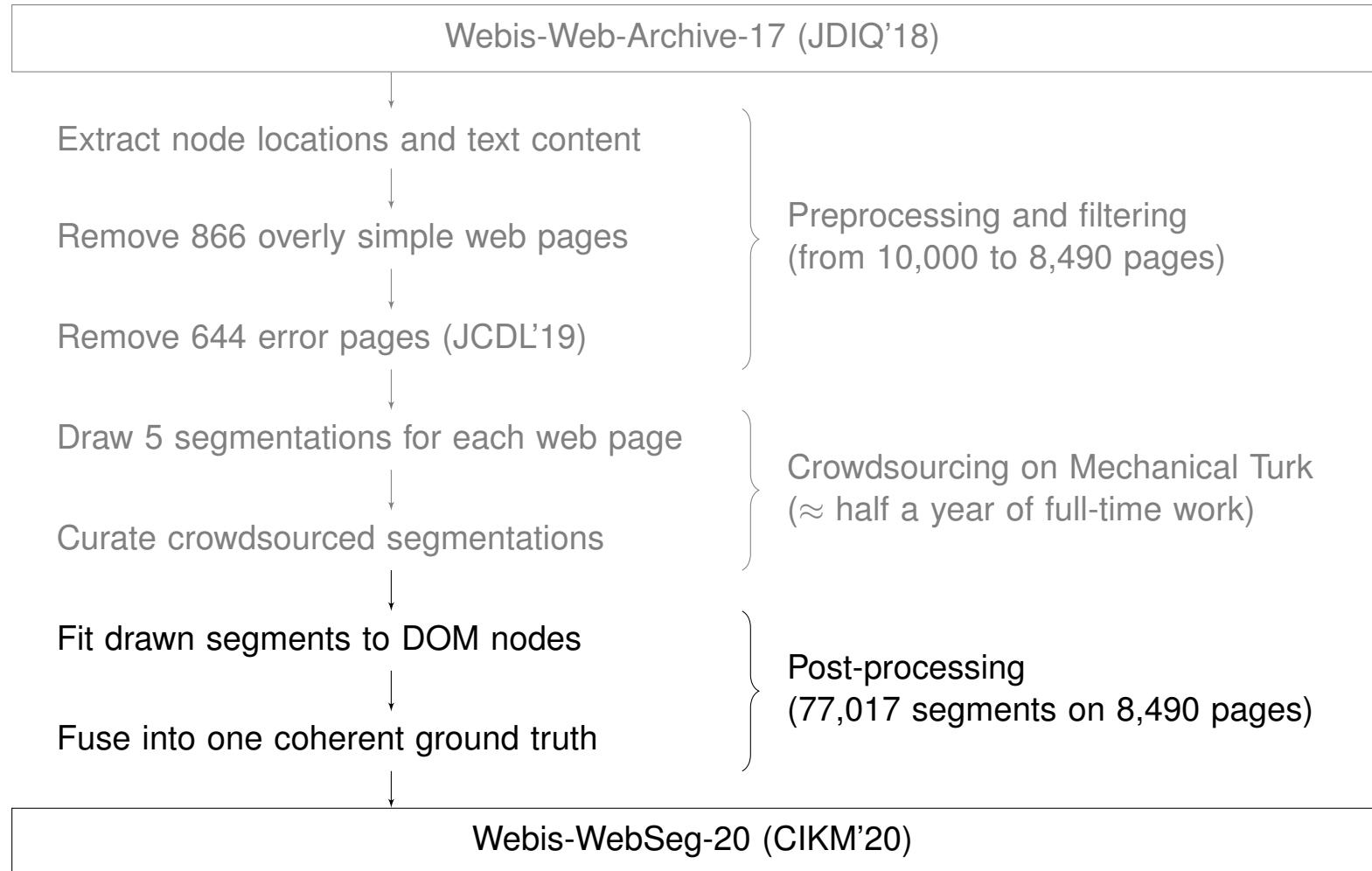
Webis-Web-Archive-17 (JDIQ'18)

Extract node locations and text content

The screenshot shows a Vietnamese news website with the following details:

- Header:** Includes a search bar, date (Thứ sáu, 22/09/2017 | 05:40 GMT+7), location (Hà Nội, 31°C), and navigation links like "Đặt làm trang chủ", "Quảng cáo", "Liên hệ", and "RSS".
- Main Navigation:** Categories include Việt Báo, Chính Trị, Thế Giới, Xã Hội, Kinh Doanh, Pháp Luật, Bóng Đá, Giải Trí, Sống Trẻ, Đời Sống, Sức Khỏe.
- Current Section:** Thé Giới Giải Trí - Chàng Nàng.
- Content:** Three news cards:
 - Bí kíp "lèo", Hoa hậu Việt Nam tung ảnh trả nãy lộ vẻ đầy đặn (22/08)
 - Bí kíp khung khép muốn chết đi của diễn viên Thành Trung (22/08)
 - 10 cặp sao Hollywood yêu nhau bất chấp tuổi tác chính thức (22/08)
- Advertisement:** A banner for a German lottery with text in German: "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." and "Klicken Sie hier".
- Article Preview:** A box titled "SỰ KIỆN NÓNG HÔM NAY" featuring a photo of a couple and the headline: "Bạn gái cặp với người nước ngoài vì tiền". It includes social sharing icons and a "Subscribe" button.
- Tags:** Vietnamese words including Việt Nam, người nước ngoài, ở nước ngoài, mối quan hệ, cô ấy, bạn gái, tình cảm, có thể, đi làm, yêu thương, tiền, tháng, nói, lời, việc.
- Text Content:** A large block of Vietnamese text describing a relationship between a Vietnamese girl and an foreign man.
- Related Content:** A sidebar titled "ĐỌC NHIỀU" with several news snippets.
- Right Side:** A large, dark, partially visible image or video frame.

The Webis-WebSeg-20 Dataset



Algorithms

Name	Reference	Document	Features	Output
VIPS	Cai et al., 2003	Web page	Tree, style, location	Rectangle tree
HEPS	Manabe and Tajima, 2015	Web page	Tree, style	Node set
Cormier et al.	Cormier et al., 2017	Web page	Screenshot	Rectangle tree
MMDetection	Chen et al., 2019	Photo	Screenshot	Pixel masks
Meier et al.	Meier et al., 2017	Article page	Screenshot, text-mask	Mask

Baseline

- One segment that covers the whole page
- Always achieves a recall of 1

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

Listen Live on iHeartRadio

On-Air News Podcasts & Media Connect Contests

Search icon

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

posted by Samantha Martin | Popular - 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.

Elton John - Your Song (Through The Years)



Photo Credit Getty Images

Facebook Twitter Email

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 Content Guidelines Advertise on News Radio 610 WTVN Get Results Contact Philanthropy © 2017 iHeartMedia, Inc.

Facebook Twitter Instagram YouTube

Listen Now

iHeartRadio

"Vision-based Page Segmentation algorithm" by Cai et al., 2003

- Starts with one segment that covers the whole page
- Computes the "degree of coherence" of each segment through heuristic rules
- Splits segments if their degree of coherence is less than the permitted degree (PDoC)

We re-implemented the algorithm to run in a modern browser

The screenshot shows a news article from iHeartRadio. The title is "Flashback: Supercut of Elton John singing 'Your Song' through the years". The article is posted by Samantha Martin | Popdust - 4 years ago. It includes a video thumbnail of Elton John singing, a photo credit to Getty Images, social sharing icons, and a comments section.

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

Listen Live on iHeartRadio

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.

Elton John - Your Song (Through The Years)

Photo Credit Getty Images

f t m +

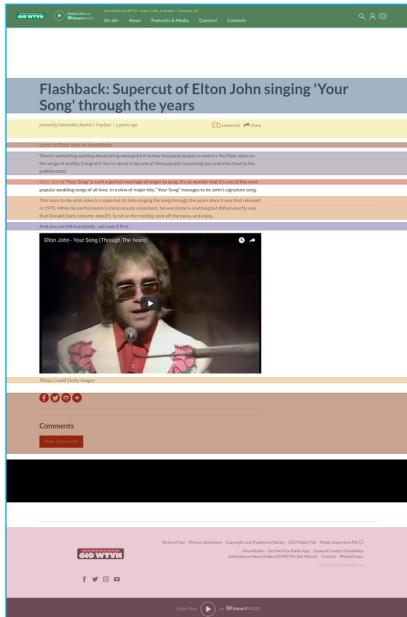
Comments

View Comments

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

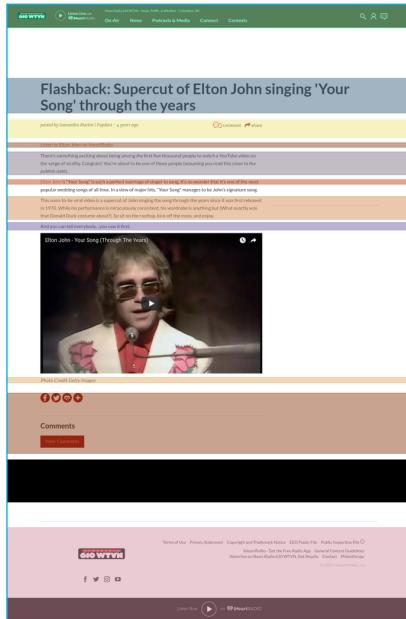
iHeartRadio Listen Now on iHeartRadio

VIPS: Optimization for Permitted Degree of Coherence (PDoC)

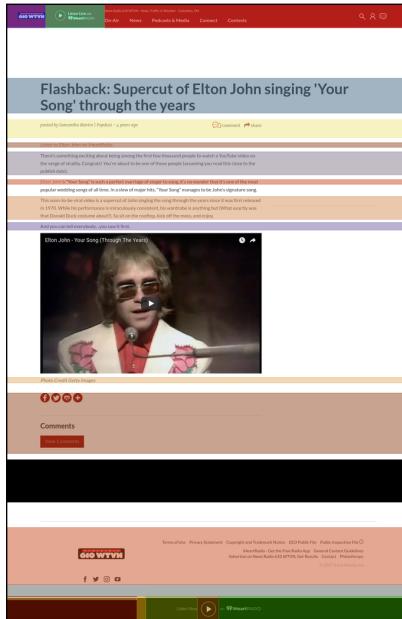


$$\text{PDoC} \in [1, 6]$$

VIPS: Optimization for Permitted Degree of Coherence (PDoC)

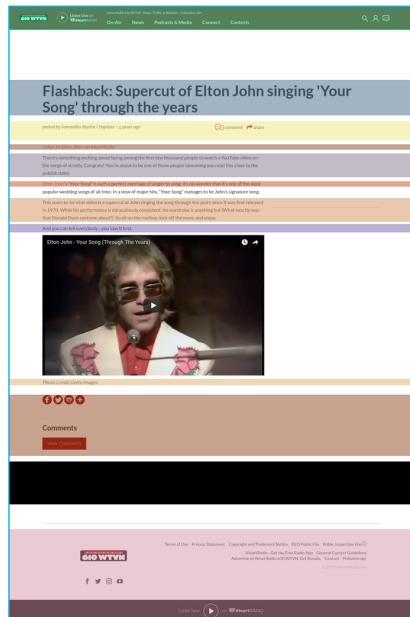


PDoC $\in [1, 6]$

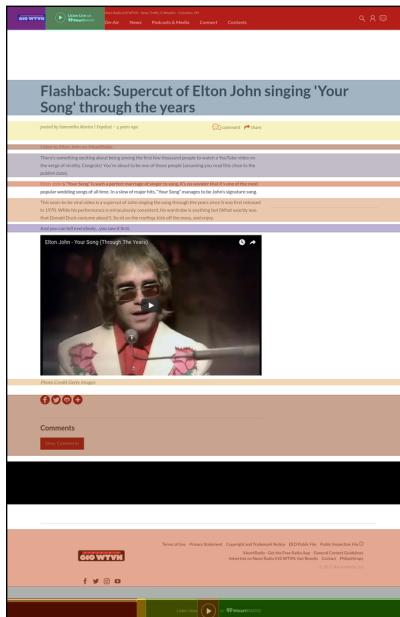


PDoC = 7

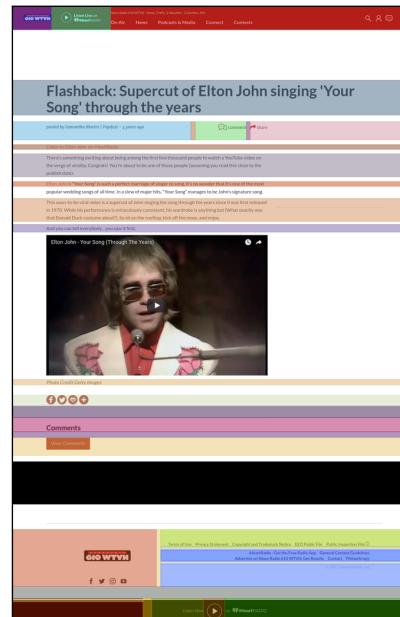
VIPS: Optimization for Permitted Degree of Coherence (PDoC)



$$\text{PDoC} \in [1, 6]$$

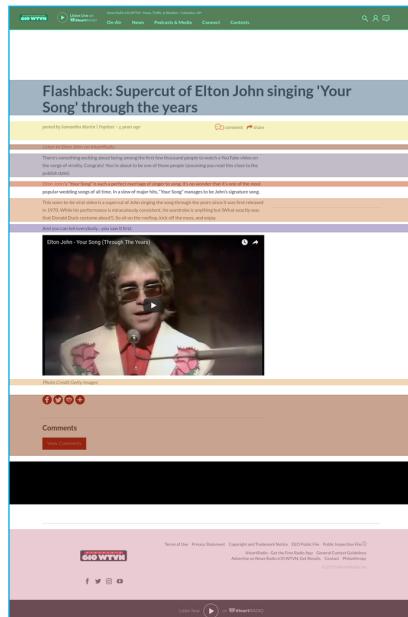


$$\text{PDoC} = 7$$

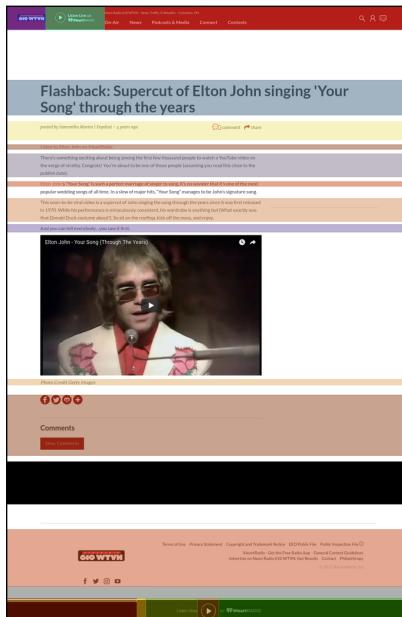


$$\text{PDoC} \in [8, 9]$$

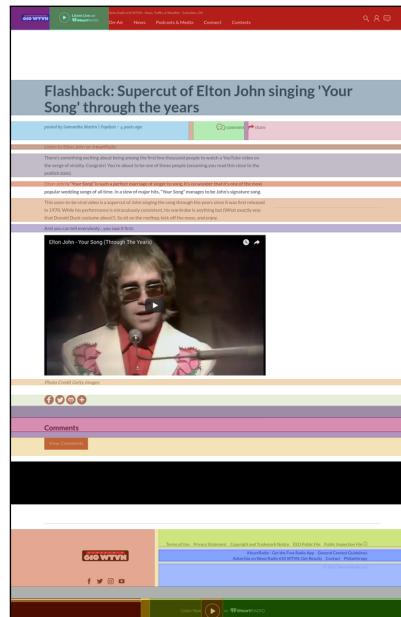
VIPS: Optimization for Permitted Degree of Coherence (PDoC)



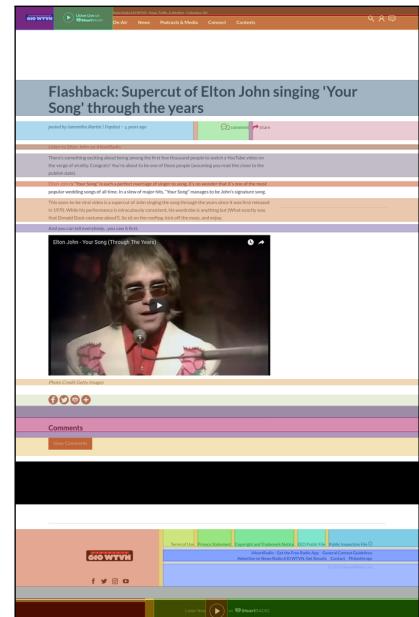
$$\text{PDoC} \in [1, 6]$$



$$\text{PDoC} = 7$$

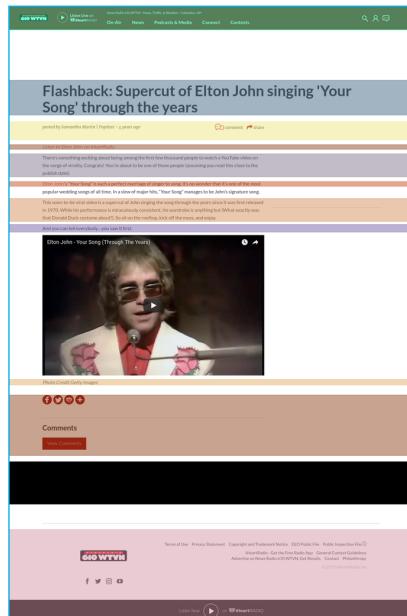


$$\text{PDoC} \in [8, 9]$$

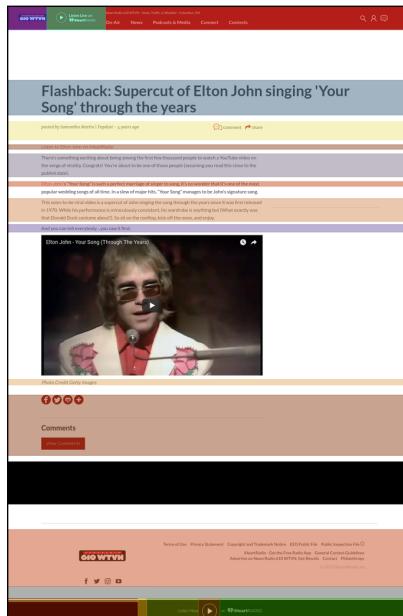


$$\text{PDoC} \in [10, 11]$$

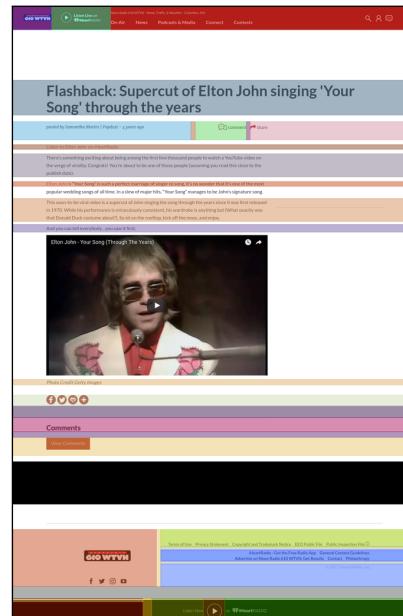
VIPS: Optimization for Permitted Degree of Coherence (PDoC)



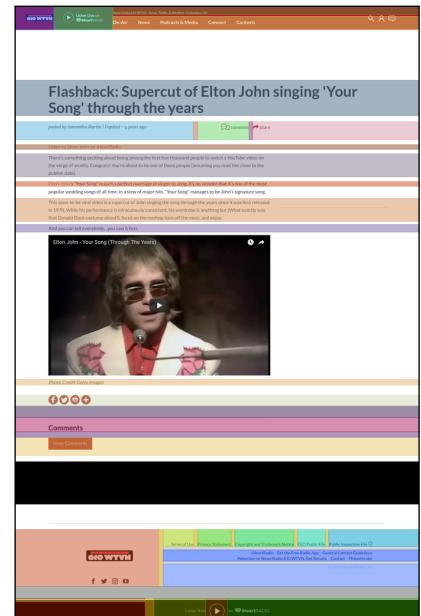
$$\text{PDoC} \in [1, 6]$$



$$\text{PDoC} = 7$$

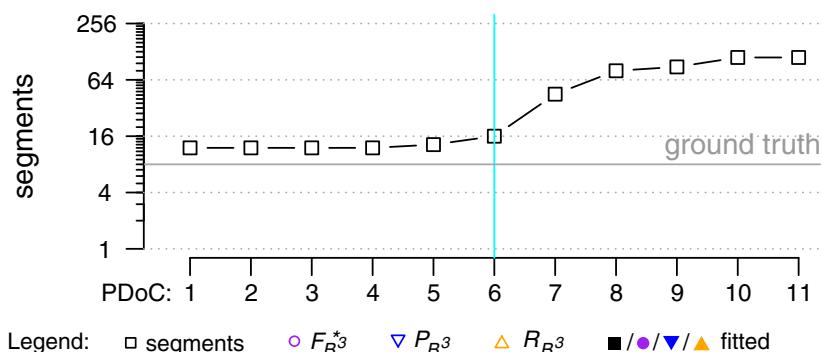


$$\text{PDoC} \in [8, 9]$$

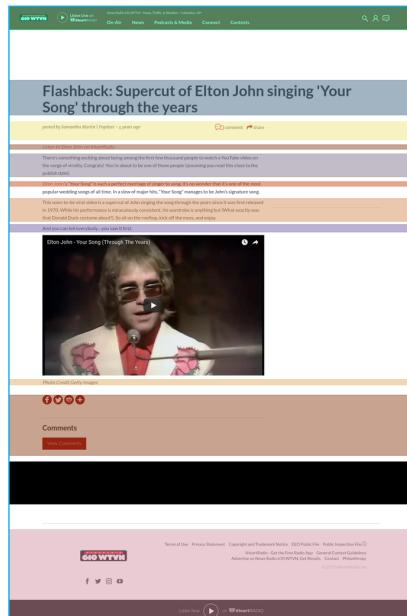


$$\text{PDoC} \in [10, 11]$$

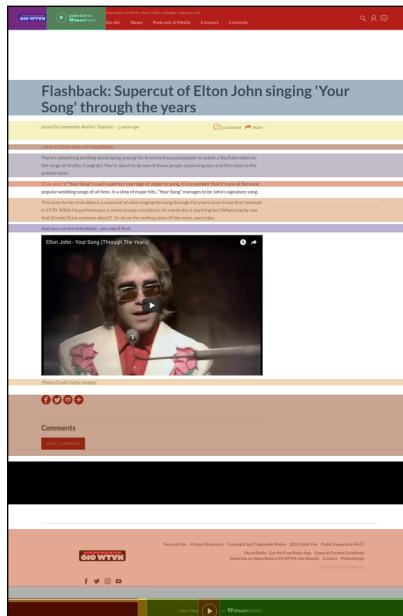
Number of segments



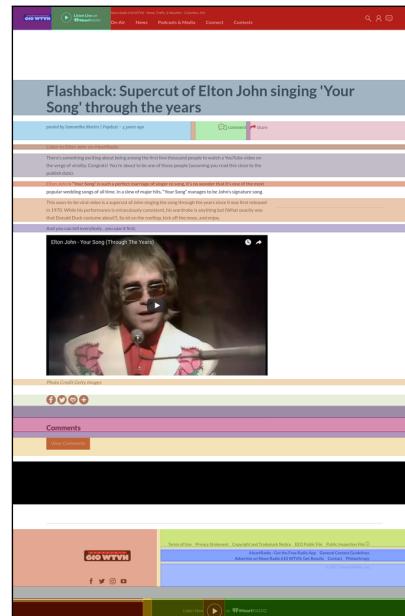
VIPS: Optimization for Permitted Degree of Coherence (PDoC)



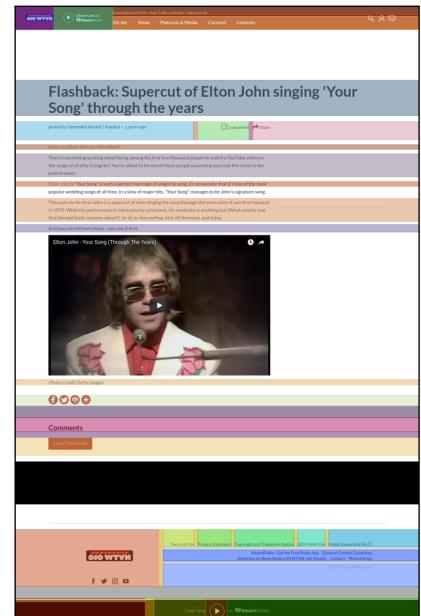
$$\text{PDoC} \in [1, 6]$$



$$\text{PDoC} = 7$$

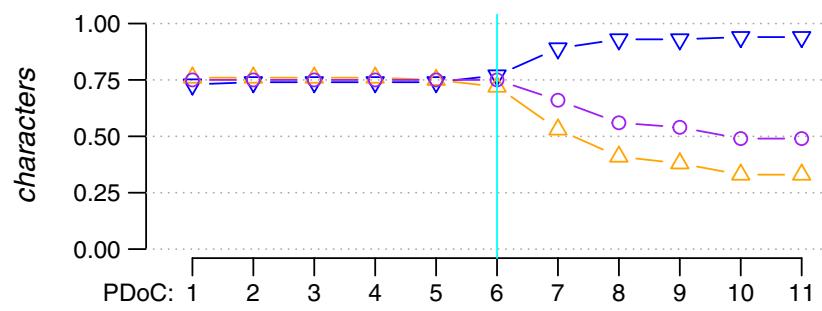
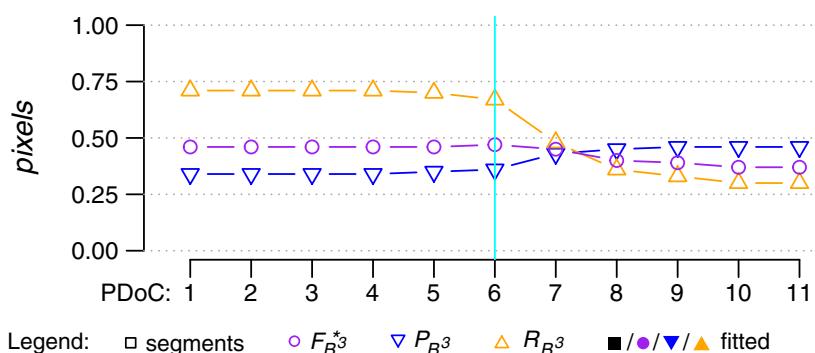


$$\text{PDoC} \in [8, 9]$$



$$\text{PDoC} \in [10, 11]$$

Comparison with ground-truth



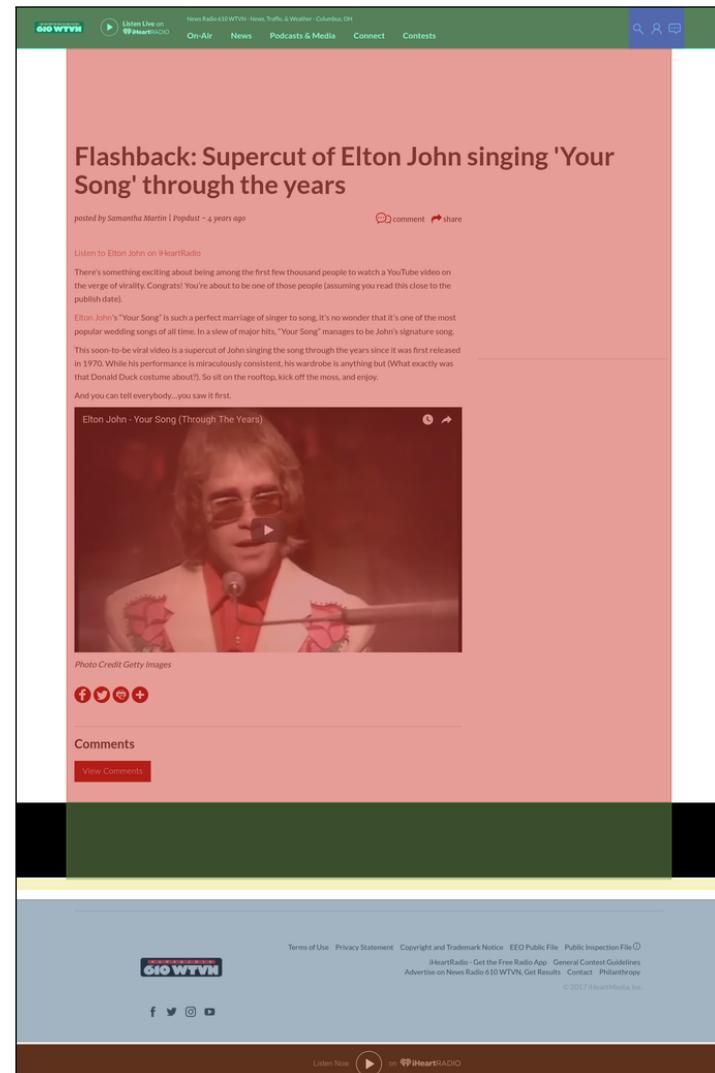
“HEading-based Page Segmentation algorithm” by Manabe and Tajima, 2015

- Identifies headings and their segments by heuristic rules
- A heading is “both visually prominent and described the topic of a segment”

We slightly adopted the author’s original implementation

The screenshot shows a news website for WTVN-TV. At the top, there's a navigation bar with links for "Live Now on iHeartRadio", "On-Air", "News", "Podcasts & Media", "Connect", and "Contests". Below the navigation, a green banner displays the headline: "Flashback: Supercut of Elton John singing 'Your Song' through the years". Underneath the headline, it says "posted by Samantha Martin | Popdust - 4 years ago" and includes "comment" and "share" buttons. The main content area features a video thumbnail of Elton John singing, with the caption "Elton John - Your Song (Through The Years)". Below the video, there's a "Photo Credit Getty Images" link and social sharing icons (Facebook, Twitter, etc.). A "Comments" section with a "View Comments" button is also visible. At the bottom of the page, there's a footer with links for "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", and copyright information: "© 2017 iHeartMedia, Inc."

- ❑ Uses the web page screenshot as sole input
- ❑ Identifies locally significant horizontal and vertical edge pixels
- ❑ Identifies horizontal and vertical “semantically significant” lines of such pixels
- ❑ Recursively splits segments by most semantically significant line



The screenshot shows the homepage of Popcash, a platform for publishers. The background features a dark blue gradient with a faint image of a city skyline at night. In the top left corner, there's a logo for "Popcash" with the subtitle "The Popunder Network". The top navigation bar includes links for "Home", "Publishers" (which is highlighted in red), "Advertisers", "FAQ", "About Us", and "Contact". On the far right of the navigation bar are "Login" and "Register" buttons. The main content area has a large white call-to-action box. Inside this box, the word "Publishers" is displayed in large white letters above a sub-headline "Maximize your revenue with PopCash.Net". Below this, there's a section titled "HOW TO START?" with a button labeled "Learn More about Our Advantages!". To the right of the main content, there's a separate sign-up form with fields for "Full Name" and "Email", and a green "Create Account" button.

Publishers

Maximize your revenue with PopCash.Net

HOW TO START?

Learn More about Our Advantages!

Sign Up Now!

Start earning money in less than 10 minutes!

Full Name

Email

Create Account

The image shows a screenshot of the Popcash website. On the left side, there is a dark blue background with a faint grid pattern. At the top left, the Popcash logo is displayed with the tagline "The Popunder Network". Below the logo, there is a graphic icon featuring a megaphone and some stars. The word "Publishers" is prominently displayed in large white letters. Below it, the text "Maximize your revenue with PopCash.Net" is shown. In the center, there is a "HOW TO START?" section with a button labeled "Learn More about Our Advantages!". On the right side, a white rectangular sign-up form is overlaid. The form has a title "Sign Up Now!" and a subtitle "Start earning money in less than 10 minutes!". It contains two input fields: "Full Name" and "Email", both represented by white input boxes with black outlines. A green button at the bottom right of the form is labeled "Create Account". At the very top of the page, there is a navigation bar with links: Home, Publishers (which is highlighted in red), Advertisers, FAQ, About Us, Contact, Login, and Register.

Popcash
The Popunder Network

+ *
Publishers

Maximize your revenue with PopCash.Net

HOW TO START?

Learn More about Our Advantages!

Sign Up Now!

Start earning money in less than 10 minutes!

Full Name

Email

Create Account

Home Publishers Advertisers FAQ About Us Contact Login Register

The image shows a screenshot of the Popcash website. On the left side, there is a dark blue background with a faint grid pattern. At the top left, the Popcash logo is displayed with the tagline "The Popunder Network". Below the logo, there is a graphic icon featuring a megaphone and some stars. The word "Publishers" is prominently displayed in large white letters. Below "Publishers", the text "Maximize your revenue with PopCash.Net" is shown. In the center, there is a white rectangular sign-up form. The form has the heading "Sign Up Now!" and the subtext "Start earning money in less than 10 minutes!". It contains two input fields: one for "Full Name" and one for "Email", both represented by white input boxes. At the bottom right of the form is a green button labeled "Create Account". At the very top of the page, there is a navigation bar with links for "Home", "Publishers" (which is highlighted in red), "Advertisers", "FAQ", "About Us", and "Contact". To the right of the navigation bar are two buttons: "Login" and "Register", with "Login" being orange and "Register" being red.

Popcash
The Popunder Network

+ *
Publishers

Publishers

Maximize your revenue with PopCash.Net

HOW TO START?

Learn More about Our Advantages!

Sign Up Now!

Start earning money in less than 10 minutes!

Full Name

Email

Create Account

Home Publishers Advertisers FAQ About Us Contact

Login Register

The image shows a screenshot of the Popcash website. On the left, there's a dark blue background with a megaphone icon and the word "Publishers". Below it, a button says "Learn More about Our Advantages!". On the right, there's a white sign-up form with fields for "Full Name" and "Email", and a green "Create Account" button. The top navigation bar includes links for Home, Publishers (which is highlighted in red), Advertisers, FAQ, About Us, Contact, Login, and Register.

Popcash
The Popunder Network

Home Publishers Advertisers FAQ About Us Contact Login Register

Sign Up Now!

Start earning money in less than 10 minutes!

Full Name

Email

Create Account

Publishers

Maximize your revenue with PopCash.Net

HOW TO START?

Learn More about Our Advantages!

The image shows a screenshot of the Popcash website. On the left side, there is a dark blue background featuring a city skyline at night. Overlaid on this are several white text elements and icons. At the top left is the "Popcash" logo with the tagline "The Popunder Network". Below the logo is a megaphone icon with a gear and a plus sign inside it. The word "Publishers" is prominently displayed in large white letters. Below "Publishers" is the text "Maximize your revenue with PopCash.Net". Further down, there is a "HOW TO START?" section with a "Learn More about Our Advantages!" button. On the right side of the image, there is a white rectangular sign-up form. The form has a heading "Sign Up Now!" and the subtext "Start earning money in less than 10 minutes!". It contains two input fields: one for "Full Name" and one for "Email", both with placeholder text. A green "Create Account" button is located at the bottom of the form. The top navigation bar of the website includes links for Home, Publishers (which is highlighted in red), Advertisers, FAQ, About Us, Contact, Login, and Register.

Popcash
The Popunder Network

+ *
Publishers

Maximize your revenue with PopCash.Net

HOW TO START?

Learn More about Our Advantages!

Home Publishers Advertisers FAQ About Us Contact Login Register

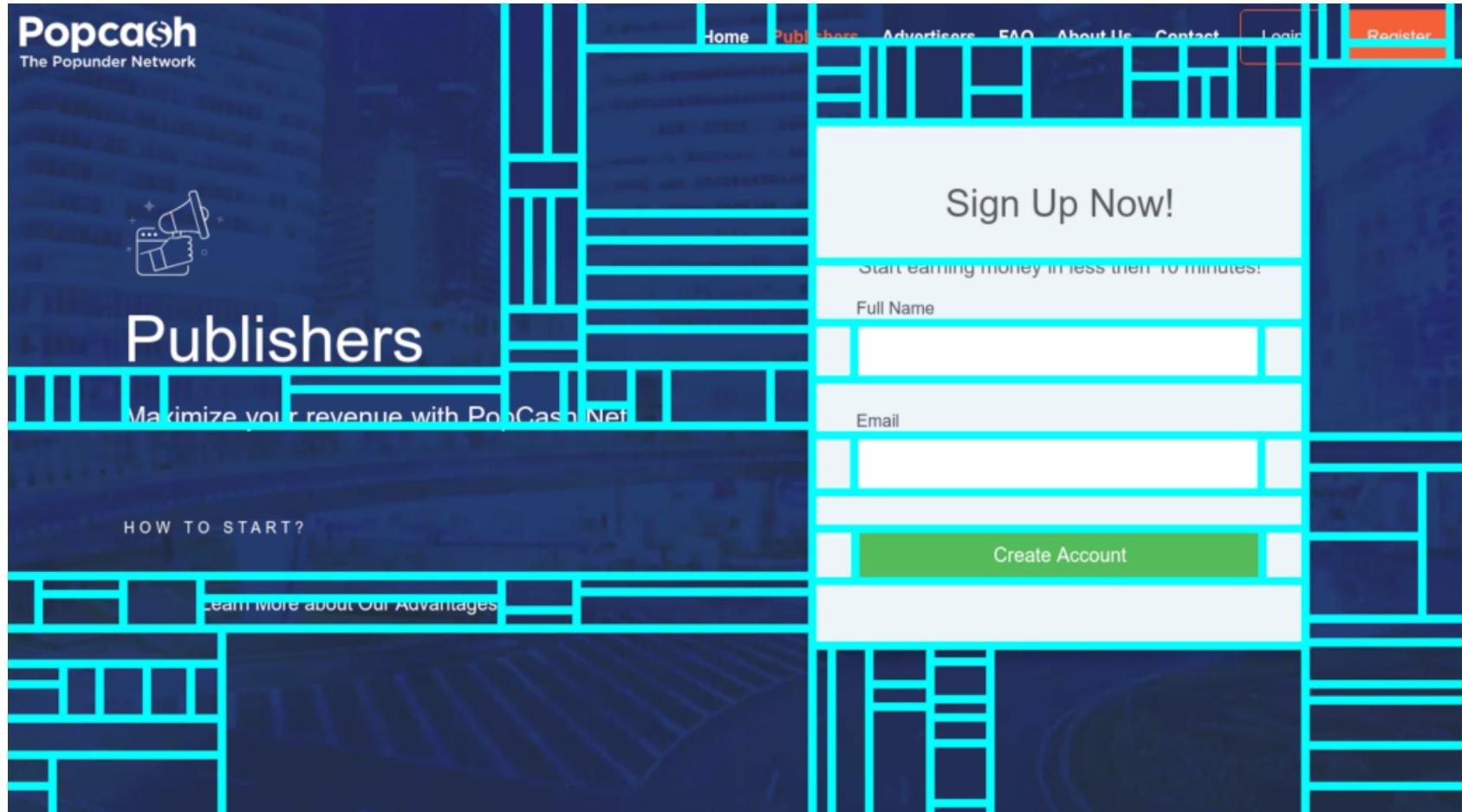
Sign Up Now!

Start earning money in less than 10 minutes!

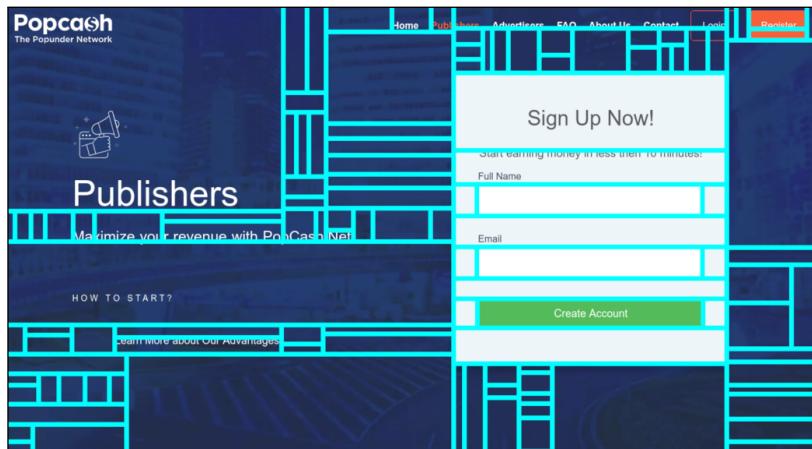
Full Name

Email

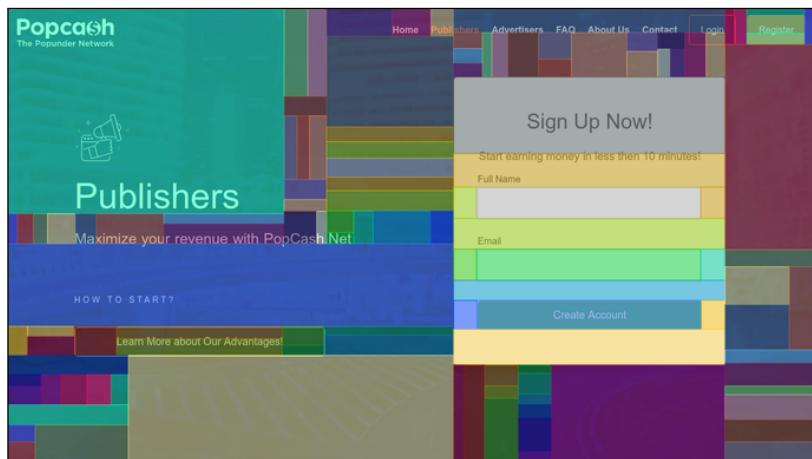
Create Account



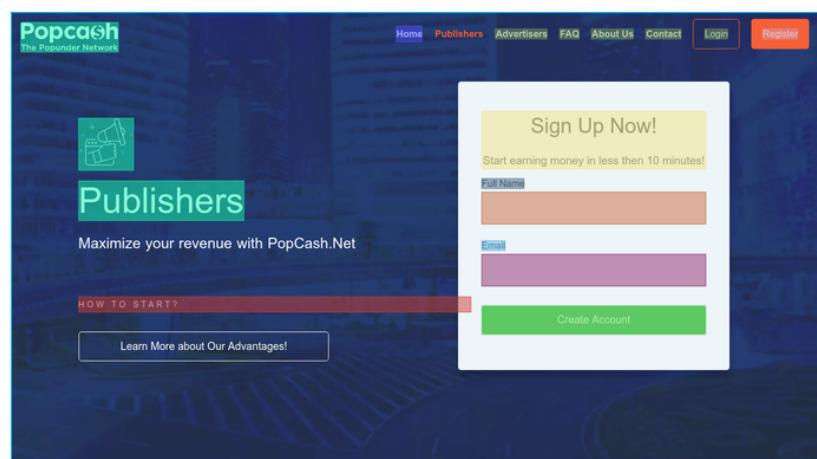
Cormier et al.: Fitting to DOM Nodes



Original (borders)

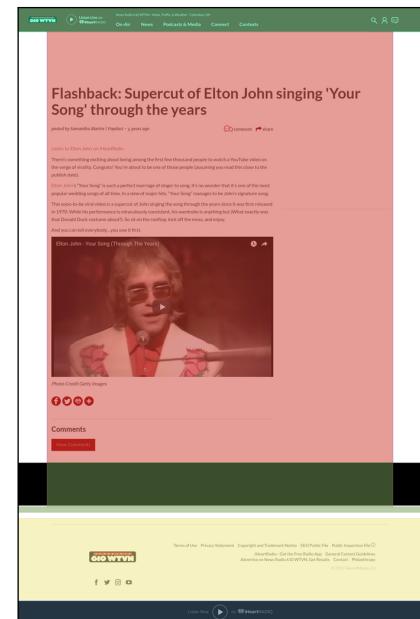
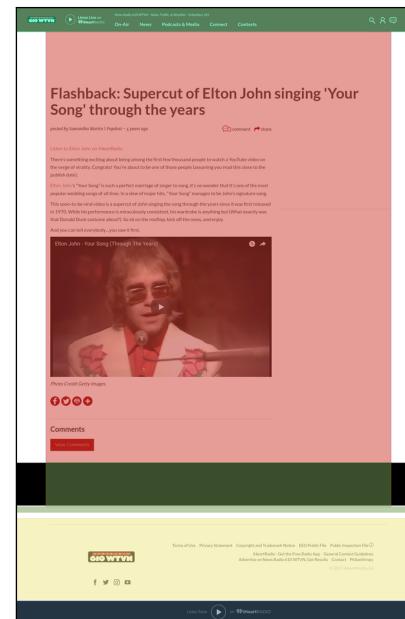
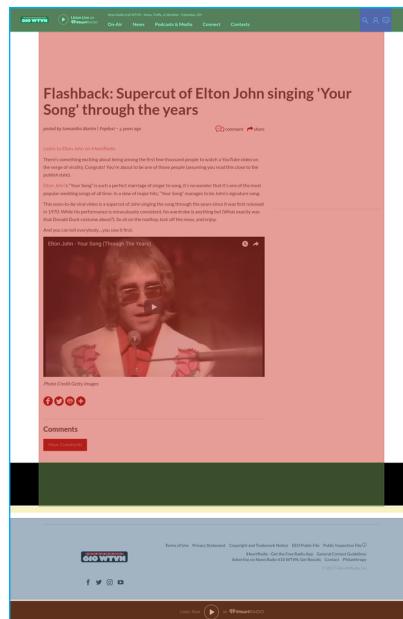
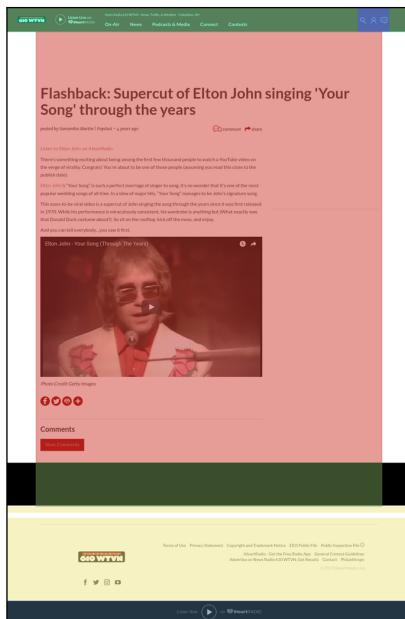


Original



Fitted

Cormier et al.: Optimization



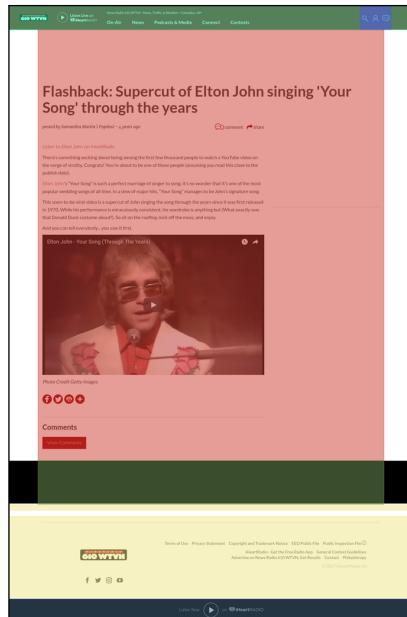
$t_l = 256; s_{\min} = 45$

$t_l = 512; s_{\min} = 45$

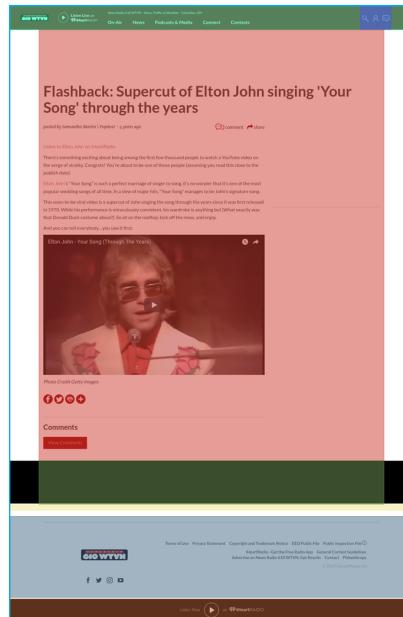
$t_l = 256; s_{\min} = 90$

$t_l = 512; s_{\min} = 90$

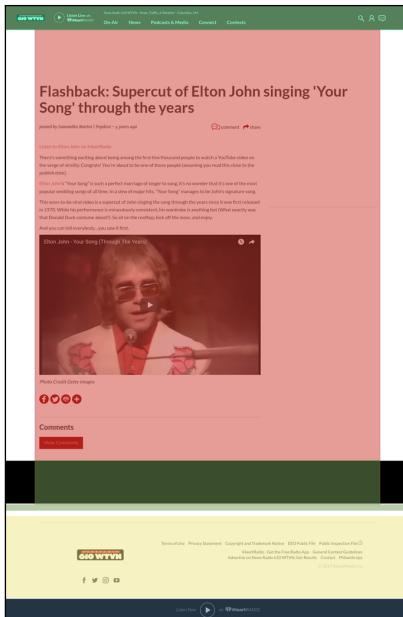
Cormier et al.: Optimization



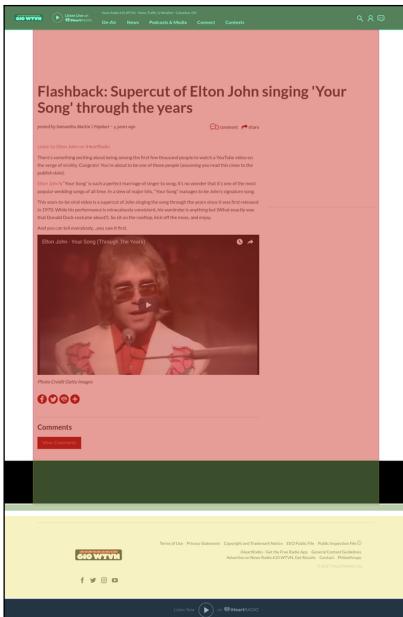
Flashback: Supcut of Elton John singing 'Your Song' through the years



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



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



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

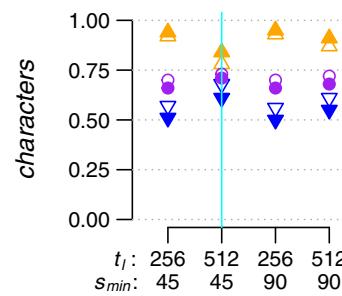
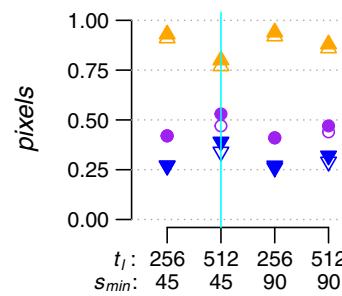
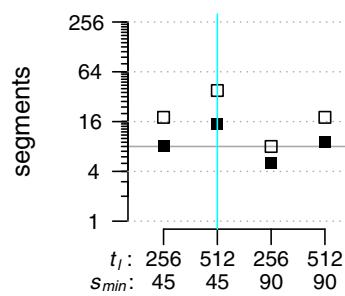
$t_l = 256; s_{\min} = 45$

$t_l = 512; s_{\min} = 45$

$t_l = 256; s_{\min} = 90$

$t_l = 512; s_{\min} = 90$

Number of segments and comparison with ground-truth



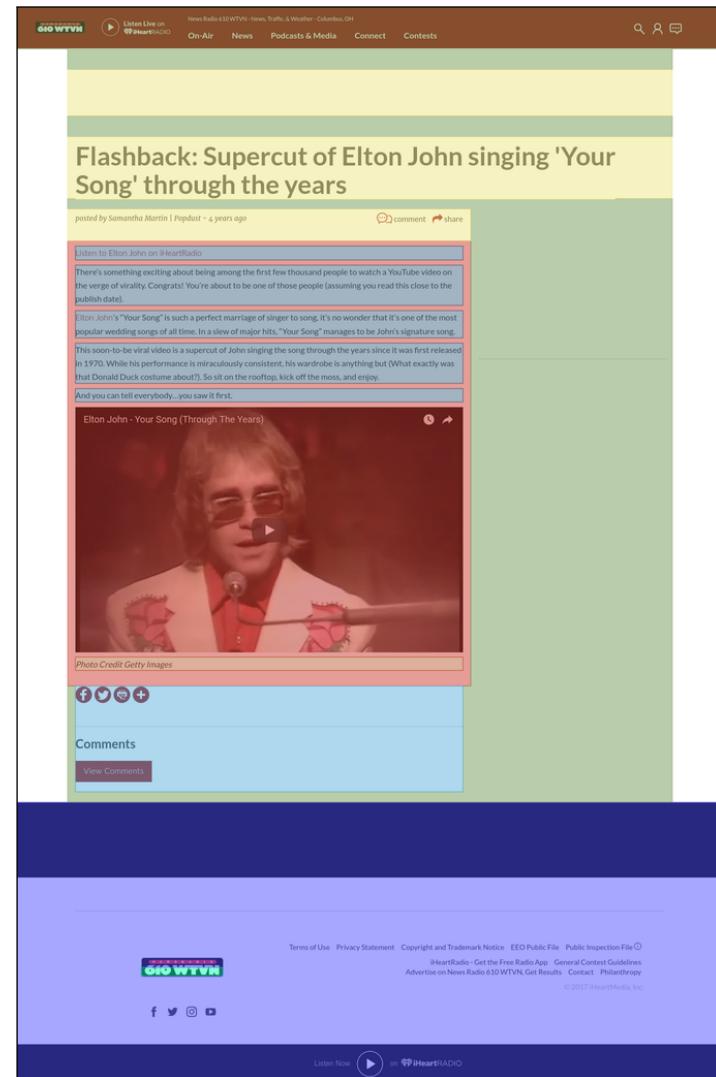
Legend: \square segments \circ F_B^3 \triangledown P_B^3 \triangle R_B^3 ■ / ● / ▼ / ▲ fitted

MMDetection

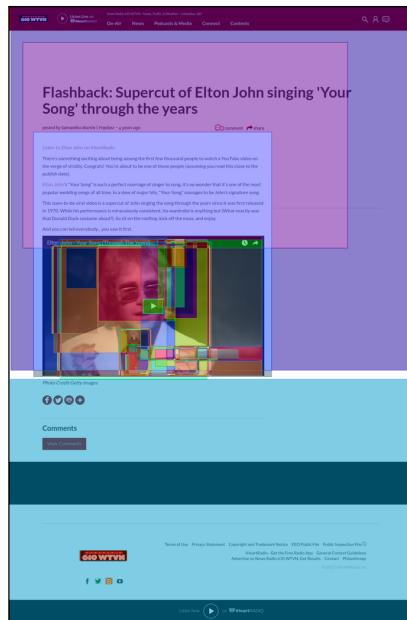
One Hybrid Task Cascade model from the MMDetection toolbox by Chen et al., 2019.

Model was state-of-the-art in 2020 as per the MSCOCO object detection task leaderboard

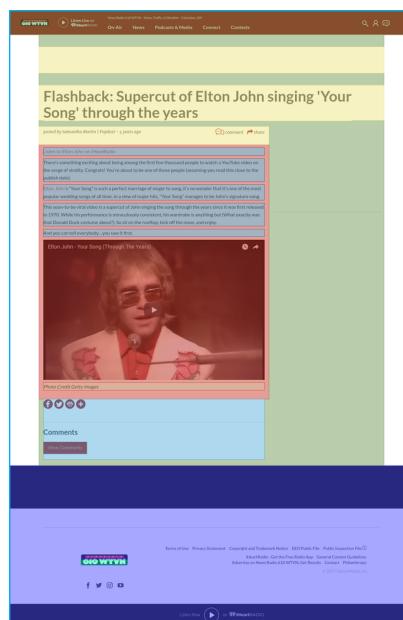
- ❑ Uses the web page screenshot as sole input
- ❑ Neural network
- ❑ Trained on object detection in real-world images (photos)



MMDetection: Optimization

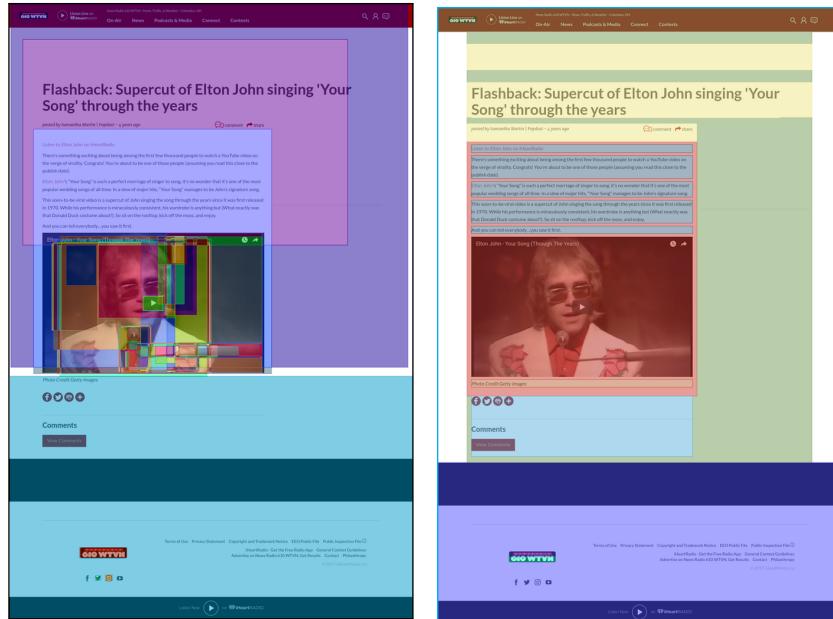


Original



Fitted

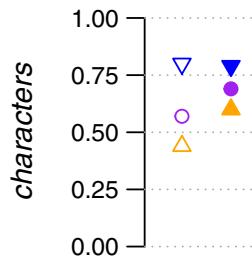
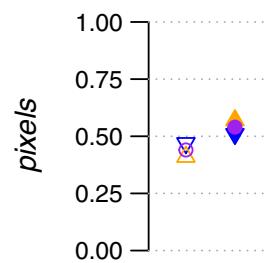
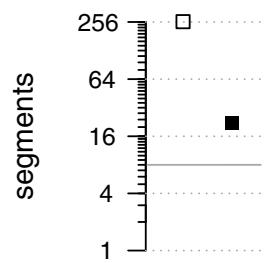
MMDetection: Optimization



Original

Fitted

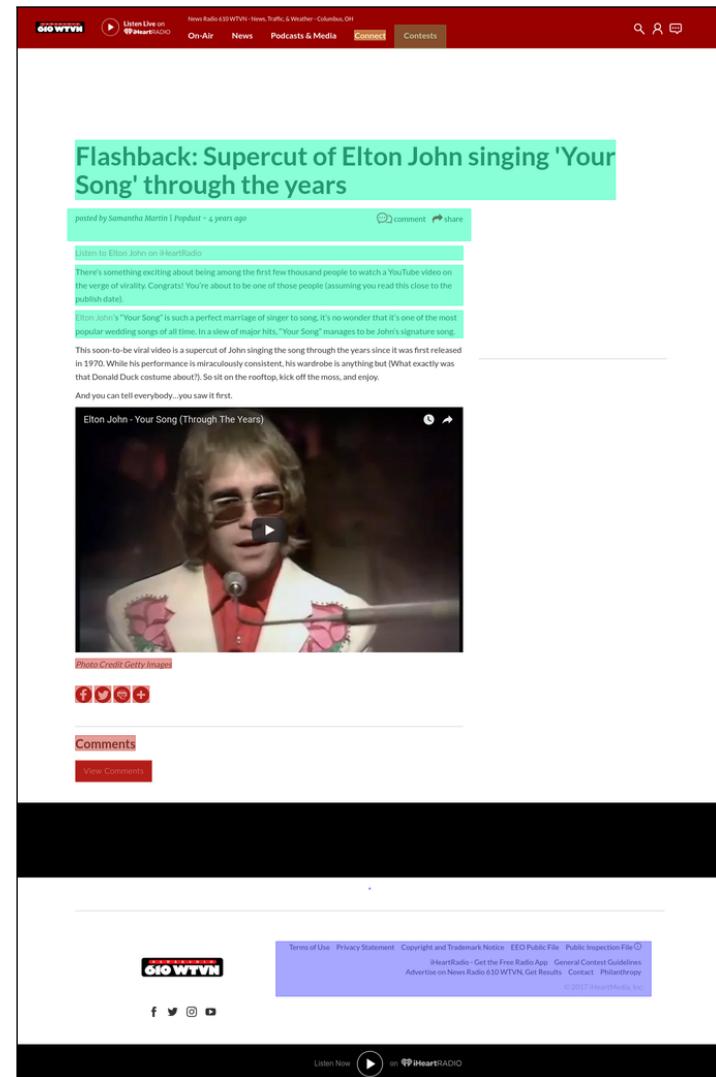
Number of segments and comparison with ground-truth



Legend: \square segments $\circ F_B^{*3}$ $\triangledown P_B^{*3}$ $\triangle R_B^{*3}$ ■ / ● / ▼ / ▲ fitted

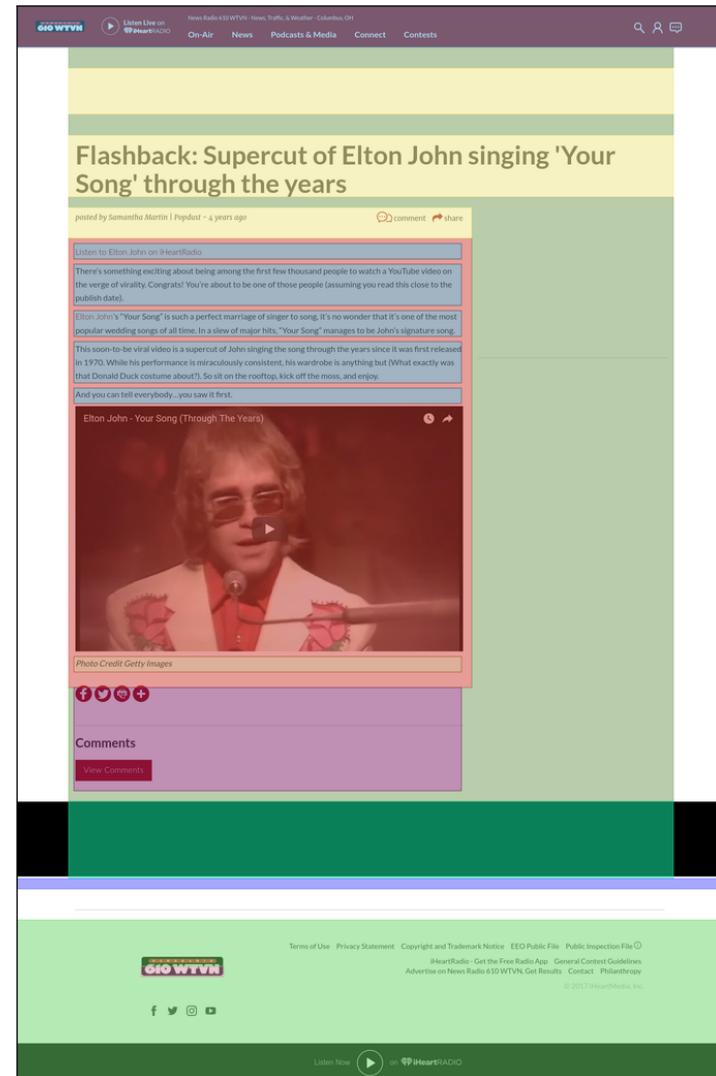
Meier et al., 2017

- ❑ Uses the web page screenshot and the location of text nodes as input
- ❑ Convolutional neural network
- ❑ Requires fixed-size input images: cropping to 4096 pixels height
- ❑ Originally developed/trained for newspaper segmentation
- ❑ 10-fold cross-evaluation on the Webis-WebSeg-20
- ❑ No detailed comparison to other algorithms due to differences in the setup

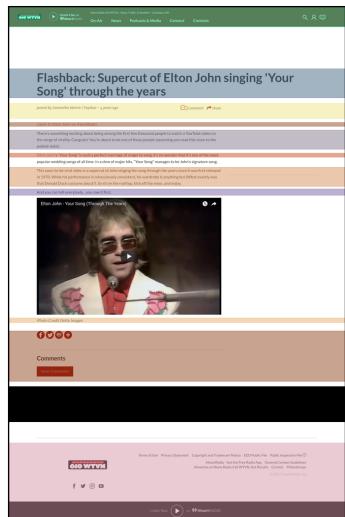


Min-vote Ensemble

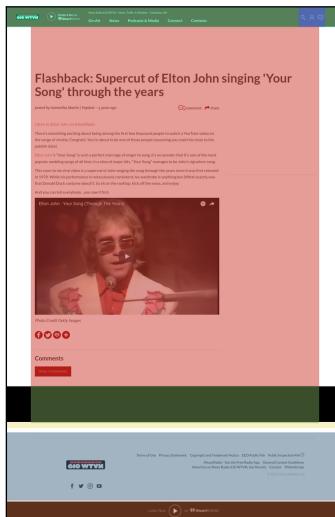
- ❑ Ensemble of VIPS, HEPS, Cormier et al., and MMDetection
- ❑ Parameter $n \in [1, 4]$
- ❑ Ignores elements which less than n algorithms placed into segments
- ❑ Standard hierarchical agglomerative clustering
- ❑ Similarity of two elements is the ratio of algorithms that place these elements in the same segment
- ❑ Similarity thresholds is $\frac{n-0.5}{4}$
Roughly: group elements together if at least n algorithms did so



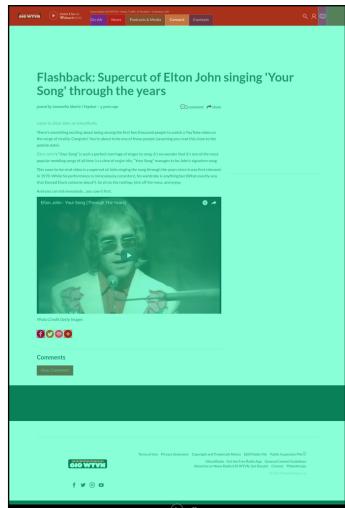
Min-vote Ensemble



VIPS



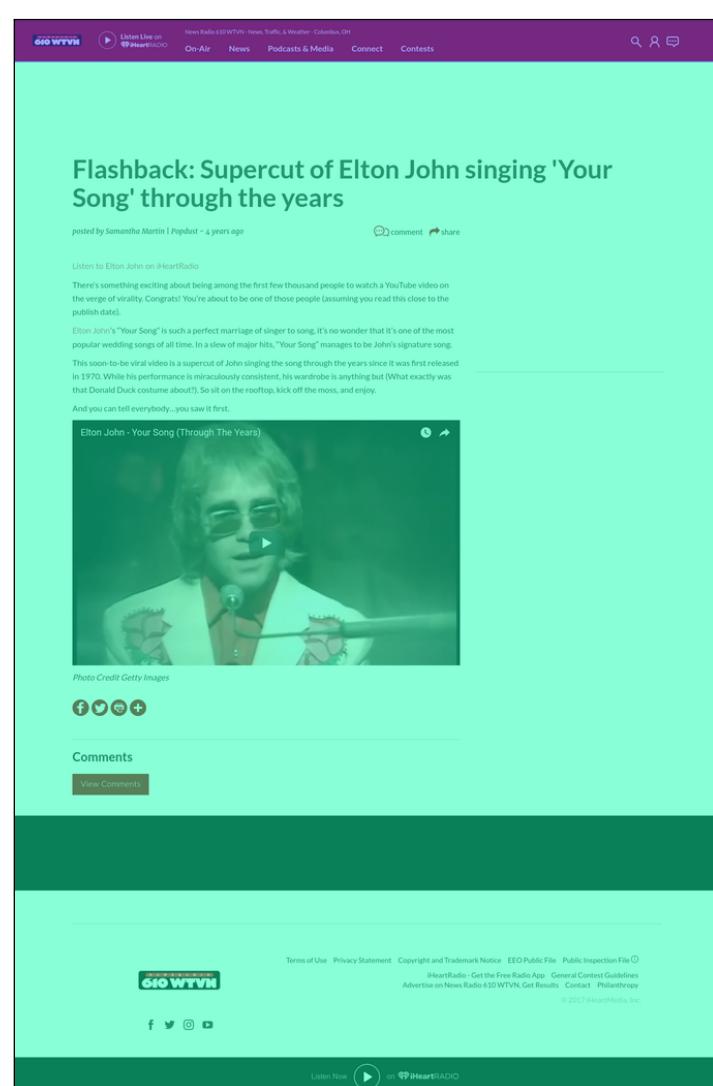
Cormier et al.



HEPS



MMDetection



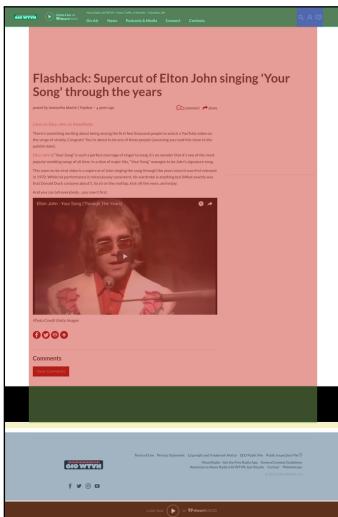
Min-vote@1

@KieselJohannes

Min-vote Ensemble



VIPS



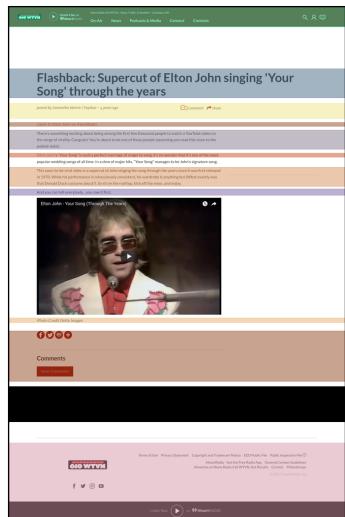
Cormier et al.



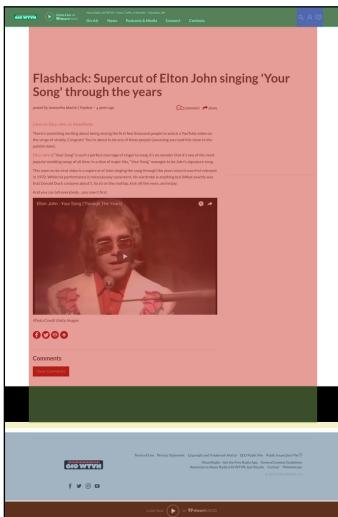
This screenshot shows a post titled "Flashback: Supercut of Elton John singing 'Your Song' through the years". It includes a video thumbnail of Elton John performing, social sharing buttons, and a comment section. A large green overlay covers the right side of the page, obscuring some of the content.

Min-vote@2

Min-vote Ensemble



VIPS



Cormier et al.



This screenshot shows the same news article from 610 WTVN, but with several visual modifications. The background is a solid purple color. The video thumbnail has a red overlay with the text "Photo Credit: Getty Images". The social sharing buttons are now blue and integrated into a green sidebar. The comments section is also part of this green sidebar. The overall design is more modern and visually distinct from the original VIPS and Cormier et al. versions.

Min-vote@4

Results

Measure	Baseline	VIPS	HEPS	Corm.	MMD.	Meier	MV@1	MV@2	MV@3	MV@4
Segments	1.0	16.1	36.1	15.3	23.0	4.6	6.5	18.7	36.5	69.5
pixels	F_{B^3}	0.24	0.38	0.33	0.36	0.42	0.32	0.30	0.39	0.30
	$F_{B^3}^*$	0.28	0.47	0.44	0.53	0.54	0.50	0.35	0.50	0.45
	P_{B^3}	0.16	0.36	0.36	0.39	0.51	0.48	0.22	0.38	0.60
	R_{B^3}	1.00	0.67	0.56	0.80	0.57	0.52	0.96	0.72	0.36
edges _F	F_{B^3}	0.44	0.59	0.48	0.51	0.53	0.41	0.50	0.56	0.39
	$F_{B^3}^*$	0.49	0.68	0.58	0.65	0.61	0.55	0.56	0.66	0.49
	P_{B^3}	0.32	0.66	0.61	0.55	0.73	0.55	0.40	0.61	0.81
	R_{B^3}	1.00	0.69	0.55	0.80	0.53	0.55	0.96	0.71	0.36
edges _C	F_{B^3}	0.45	0.61	0.49	0.53	0.54	0.42	0.51	0.57	0.39
	$F_{B^3}^*$	0.49	0.68	0.59	0.66	0.62	0.56	0.56	0.67	0.50
	P_{B^3}	0.32	0.67	0.62	0.56	0.74	0.55	0.40	0.63	0.82
	R_{B^3}	1.00	0.70	0.56	0.80	0.53	0.57	0.96	0.72	0.36
nodes	F_{B^3}	0.42	0.63	0.43	0.52	0.52	0.44	0.49	0.54	0.34
	$F_{B^3}^*$	0.46	0.70	0.54	0.65	0.61	0.56	0.55	0.65	0.44
	P_{B^3}	0.30	0.69	0.63	0.53	0.74	0.52	0.38	0.64	0.85
	R_{B^3}	1.00	0.71	0.46	0.82	0.51	0.61	0.96	0.65	0.29
chars	F_{B^3}	0.52	0.67	0.50	0.61	0.61	0.50	0.59	0.62	0.40
	$F_{B^3}^*$	0.57	0.75	0.60	0.71	0.69	0.61	0.64	0.71	0.50
	P_{B^3}	0.39	0.77	0.73	0.61	0.79	0.59	0.48	0.72	0.90
	R_{B^3}	1.00	0.72	0.51	0.84	0.60	0.63	0.96	0.71	0.35

Results

Measure	Baseline	VIPS	HEPS	Corm.	MMD.	Meier	MV@1	MV@2	MV@3	MV@4
Segments	1.0	16.1	36.1	15.3	23.0	4.6	6.5	18.7	36.5	69.5
pixels	F_{B^3}	0.24	0.38	0.33	0.36	0.42	0.32	0.30	0.39	0.30
	$F_{B^3}^*$	0.28	0.47	0.44	0.53	0.54	0.50	0.35	0.50	0.45
	P_{B^3}	0.16	0.36	0.36	0.39	0.51	0.48	0.22	0.38	0.60
	R_{B^3}	1.00	0.67	0.56	0.80	0.57	0.52	0.96	0.72	0.36
edges _F	F_{B^3}	0.44	0.59	0.48	0.51	0.53	0.41	0.50	0.56	0.39
	$F_{B^3}^*$	0.49	0.68	0.58	0.65	0.61	0.55	0.56	0.66	0.49
	P_{B^3}	0.32	0.66	0.61	0.55	0.73	0.55	0.40	0.61	0.81
	R_{B^3}	1.00	0.69	0.55	0.80	0.53	0.55	0.96	0.71	0.36
edges _C	F_{B^3}	0.45	0.61	0.49	0.53	0.54	0.42	0.51	0.57	0.39
	$F_{B^3}^*$	0.49	0.68	0.59	0.66	0.62	0.56	0.56	0.67	0.50
	P_{B^3}	0.32	0.67	0.62	0.56	0.74	0.55	0.40	0.63	0.82
	R_{B^3}	1.00	0.70	0.56	0.80	0.53	0.57	0.96	0.72	0.36
nodes	F_{B^3}	0.42	0.63	0.43	0.52	0.52	0.44	0.49	0.54	0.34
	$F_{B^3}^*$	0.46	0.70	0.54	0.65	0.61	0.56	0.55	0.65	0.44
	P_{B^3}	0.30	0.69	0.63	0.53	0.74	0.52	0.38	0.64	0.85
	R_{B^3}	1.00	0.71	0.46	0.82	0.51	0.61	0.96	0.65	0.29
chars	F_{B^3}	0.52	0.67	0.50	0.61	0.61	0.50	0.59	0.62	0.40
	$F_{B^3}^*$	0.57	0.75	0.60	0.71	0.69	0.61	0.64	0.71	0.50
	P_{B^3}	0.39	0.77	0.73	0.61	0.79	0.59	0.48	0.72	0.90
	R_{B^3}	1.00	0.72	0.51	0.84	0.60	0.63	0.96	0.71	0.35

Results

Measure	Baseline	VIPS	HEPS	Corm.	MMD.	Meier	MV@1	MV@2	MV@3	MV@4
Segments	1.0	16.1	36.1	15.3	23.0	4.6	6.5	18.7	36.5	69.5
pixels	F_{B^3}	0.24	0.38	0.33	0.36	0.42	0.32	0.30	0.39	0.30
	$F_{B^3}^*$	0.28	0.47	0.44	0.53	0.54	0.50	0.35	0.50	0.45
	P_{B^3}	0.16	0.36	0.36	0.39	0.51	0.48	0.22	0.38	0.60
	R_{B^3}	1.00	0.67	0.56	0.80	0.57	0.52	0.96	0.72	0.36
chars	F_{B^3}	0.52	0.67	0.50	0.61	0.61	0.50	0.59	0.62	0.40
	$F_{B^3}^*$	0.57	0.75	0.60	0.71	0.69	0.61	0.64	0.71	0.50
	P_{B^3}	0.39	0.77	0.73	0.61	0.79	0.59	0.48	0.72	0.90
	R_{B^3}	1.00	0.72	0.51	0.84	0.60	0.63	0.96	0.71	0.35

Results

Measure	Baseline	VIPS	HEPS	Corm.	MMD.	Meier	MV@1	MV@2	MV@3	MV@4
Segments	1.0	16.1	36.1	15.3	23.0	4.6	6.5	18.7	36.5	69.5
<i>pixels</i>	F_{B^3}	0.24	0.38	0.33	0.36	0.42	0.32	0.30	0.39	0.30
	$F_{B^3}^*$	0.28	0.47	0.44	0.53	0.54	0.50	0.35	0.50	0.45
	P_{B^3}	0.16	0.36	0.36	0.39	0.51	0.48	0.22	0.38	0.60
	R_{B^3}	1.00	0.67	0.56	0.80	0.57	0.52	0.96	0.72	0.36
<i>chars</i>	F_{B^3}	0.52	0.67	0.50	0.61	0.61	0.50	0.59	0.62	0.40
	$F_{B^3}^*$	0.57	0.75	0.60	0.71	0.69	0.61	0.64	0.71	0.50
	P_{B^3}	0.39	0.77	0.73	0.61	0.79	0.59	0.48	0.72	0.90
	R_{B^3}	1.00	0.72	0.51	0.84	0.60	0.63	0.96	0.71	0.35

Results

Measure	Baseline	VIPS	HEPS	Corm.	MMD.	Meier	MV@1	MV@2	MV@3	MV@4
Segments	1.0	16.1	36.1	15.3	23.0	4.6	6.5	18.7	36.5	69.5
<i>pixels</i>	F_{B^3}	0.24	0.38	0.33	0.36	0.42	0.32	0.30	0.39	0.30
	$F_{B^3}^*$	0.28	0.47	0.44	0.53	0.54	0.50	0.35	0.50	0.45
	P_{B^3}	0.16	0.36	0.36	0.39	0.51	0.48	0.22	0.38	0.60
	R_{B^3}	1.00	0.67	0.56	0.80	0.57	0.52	0.96	0.72	0.36
<i>chars</i>	F_{B^3}	0.52	0.67	0.50	0.61	0.61	0.50	0.59	0.62	0.40
	$F_{B^3}^*$	0.57	0.75	0.60	0.71	0.69	0.61	0.64	0.71	0.50
	P_{B^3}	0.39	0.77	0.73	0.61	0.79	0.59	0.48	0.72	0.90
	R_{B^3}	1.00	0.72	0.51	0.84	0.60	0.63	0.96	0.71	0.35

Results

Measure	Baseline	VIPS	HEPS	Corm.	MMD.	Meier	MV@1	MV@2	MV@3	MV@4
Segments	1.0	16.1	36.1	15.3	23.0	4.6	6.5	18.7	36.5	69.5
<i>pixels</i>	F_{B^3}	0.24	0.38	0.33	0.36	0.42	0.32	0.30	0.39	0.30
	$F_{B^3}^*$	0.28	0.47	0.44	0.53	0.54	0.50	0.35	0.50	0.45
	P_{B^3}	0.16	0.36	0.36	0.39	0.51	0.48	0.22	0.38	0.60
	R_{B^3}	1.00	0.67	0.56	0.80	0.57	0.52	0.96	0.72	0.36
<i>chars</i>	F_{B^3}	0.52	0.67	0.50	0.61	0.61	0.50	0.59	0.62	0.40
	$F_{B^3}^*$	0.57	0.75	0.60	0.71	0.69	0.61	0.64	0.71	0.50
	P_{B^3}	0.39	0.77	0.73	0.61	0.79	0.59	0.48	0.72	0.90
	R_{B^3}	1.00	0.72	0.51	0.84	0.60	0.63	0.96	0.71	0.35

Results

Measure	Baseline	VIPS	HEPS	Corm.	MMD.	Meier	MV@1	MV@2	MV@3	MV@4
Segments	1.0	16.1	36.1	15.3	23.0	4.6	6.5	18.7	36.5	69.5
pixels	F_{B^3}	0.24	0.38	0.33	0.36	0.42	0.32	0.30	0.39	0.30
	$F_{B^3}^*$	0.28	0.47	0.44	0.53	0.54	0.50	0.35	0.50	0.45
	P_{B^3}	0.16	0.36	0.36	0.39	0.51	0.48	0.22	0.38	0.60
	R_{B^3}	1.00	0.67	0.56	0.80	0.57	0.52	0.96	0.72	0.36
chars	F_{B^3}	0.52	0.67	0.50	0.61	0.61	0.50	0.59	0.62	0.40
	$F_{B^3}^*$	0.57	0.75	0.60	0.71	0.69	0.61	0.64	0.71	0.50
	P_{B^3}	0.39	0.77	0.73	0.61	0.79	0.59	0.48	0.72	0.90
	R_{B^3}	1.00	0.72	0.51	0.84	0.60	0.63	0.96	0.71	0.35

Results

Measure	Baseline	VIPS	HEPS	Corm.	MMD.	Meier	MV@1	MV@2	MV@3	MV@4
Segments	1.0	16.1	36.1	15.3	23.0	4.6	6.5	18.7	36.5	69.5
pixels	F_{B^3}	0.24	0.38	0.33	0.36	0.42	0.32	0.30	0.39	0.30
	$F_{B^3}^*$	0.28	0.47	0.44	0.53	0.54	0.50	0.35	0.50	0.45
	P_{B^3}	0.16	0.36	0.36	0.39	0.51	0.48	0.22	0.38	0.60
	R_{B^3}	1.00	0.67	0.56	0.80	0.57	0.52	0.96	0.72	0.36
chars	F_{B^3}	0.52	0.67	0.50	0.61	0.61	0.50	0.59	0.62	0.40
	$F_{B^3}^*$	0.57	0.75	0.60	0.71	0.69	0.61	0.64	0.71	0.50
	P_{B^3}	0.39	0.77	0.73	0.61	0.79	0.59	0.48	0.72	0.90
	R_{B^3}	1.00	0.72	0.51	0.84	0.60	0.63	0.96	0.71	0.35

Results

Measure	Baseline	VIPS	HEPS	Corm.	MMD.	Meier	MV@1	MV@2	MV@3	MV@4
Segments	1.0	16.1	36.1	15.3	23.0	4.6	6.5	18.7	36.5	69.5
pixels	F_{B^3}	0.24	0.38	0.33	0.36	0.42	0.32	0.30	0.39	0.30
	$F_{B^3}^*$	0.28	0.47	0.44	0.53	0.54	0.50	0.35	0.50	0.45
	P_{B^3}	0.16	0.36	0.36	0.39	0.51	0.48	0.22	0.38	0.60
	R_{B^3}	1.00	0.67	0.56	0.80	0.57	0.52	0.96	0.72	0.36
chars	F_{B^3}	0.52	0.67	0.50	0.61	0.61	0.50	0.59	0.62	0.40
	$F_{B^3}^*$	0.57	0.75	0.60	0.71	0.69	0.61	0.64	0.71	0.50
	P_{B^3}	0.39	0.77	0.73	0.61	0.79	0.59	0.48	0.72	0.90
	R_{B^3}	1.00	0.72	0.51	0.84	0.60	0.63	0.96	0.71	0.35

Results

Measure	Baseline	VIPS	HEPS	Corm.	MMD.	Meier	MV@1	MV@2	MV@3	MV@4
Segments	1.0	16.1	36.1	15.3	23.0	4.6	6.5	18.7	36.5	69.5
pixels	F_{B^3}	0.24	0.38	0.33	0.36	0.42	0.32	0.30	0.39	0.30
	$F_{B^3}^*$	0.28	0.47	0.44	0.53	0.54	0.50	0.35	0.50	0.45
	P_{B^3}	0.16	0.36	0.36	0.39	0.51	0.48	0.22	0.38	0.60
	R_{B^3}	1.00	0.67	0.56	0.80	0.57	0.52	0.96	0.72	0.36
chars	F_{B^3}	0.52	0.67	0.50	0.61	0.61	0.50	0.59	0.62	0.40
	$F_{B^3}^*$	0.57	0.75	0.60	0.71	0.69	0.61	0.64	0.71	0.50
	P_{B^3}	0.39	0.77	0.73	0.61	0.79	0.59	0.48	0.72	0.90
	R_{B^3}	1.00	0.72	0.51	0.84	0.60	0.63	0.96	0.71	0.35

Results

Measure	Baseline	VIPS	HEPS	Corm.	MMD.	Meier	MV@1	MV@2	MV@3	MV@4
Segments	1.0	16.1	36.1	15.3	23.0	4.6	6.5	18.7	36.5	69.5
<i>pixels</i>	F_{B^3}	0.24	0.38	0.33	0.36	0.42	0.32	0.30	0.39	0.30
	$F_{B^3}^*$	0.28	0.47	0.44	0.53	0.54	0.50	0.35	0.50	0.45
	P_{B^3}	0.16	0.36	0.36	0.39	0.51	0.48	0.22	0.38	0.60
	R_{B^3}	1.00	0.67	0.56	0.80	0.57	0.52	0.96	0.72	0.36
<i>chars</i>	F_{B^3}	0.52	0.67	0.50	0.61	0.61	0.50	0.59	0.62	0.40
	$F_{B^3}^*$	0.57	0.75	0.60	0.71	0.69	0.61	0.64	0.71	0.50
	P_{B^3}	0.39	0.77	0.73	0.61	0.79	0.59	0.48	0.72	0.90
	R_{B^3}	1.00	0.72	0.51	0.84	0.60	0.63	0.96	0.71	0.35

Results

Measure	Baseline	VIPS	HEPS	Corm.	MMD.	Meier	MV@1	MV@2	MV@3	MV@4
Segments	1.0	16.1	36.1	15.3	23.0	4.6	6.5	18.7	36.5	69.5
<i>pixels</i>	F_{B^3}	0.24	0.38	0.33	0.36	0.42	0.32	0.30	0.39	0.30
	$F_{B^3}^*$	0.28	0.47	0.44	0.53	0.54	0.50	0.35	0.50	0.45
	P_{B^3}	0.16	0.36	0.36	0.39	0.51	0.48	0.22	0.38	0.60
	R_{B^3}	1.00	0.67	0.56	0.80	0.57	0.52	0.96	0.72	0.36
<i>chars</i>	F_{B^3}	0.52	0.67	0.50	0.61	0.61	0.50	0.59	0.62	0.40
	$F_{B^3}^*$	0.57	0.75	0.60	0.71	0.69	0.61	0.64	0.71	0.50
	P_{B^3}	0.39	0.77	0.73	0.61	0.79	0.59	0.48	0.72	0.90
	R_{B^3}	1.00	0.72	0.51	0.84	0.60	0.63	0.96	0.71	0.35

Results

Measure	Baseline	VIPS	HEPS	Corm.	MMD.	Meier	MV@1	MV@2	MV@3	MV@4
Segments	1.0	16.1	36.1	15.3	23.0	4.6	6.5	18.7	36.5	69.5
<i>pixels</i>	F_{B^3}	0.24	0.38	0.33	0.36	0.42	0.32	0.30	0.39	0.30
	$F_{B^3}^*$	0.28	0.47	0.44	0.53	0.54	0.50	0.35	0.50	0.45
	P_{B^3}	0.16	0.36	0.36	0.39	0.51	0.48	0.22	0.38	0.60
	R_{B^3}	1.00	0.67	0.56	0.80	0.57	0.52	0.96	0.72	0.36
<i>chars</i>	F_{B^3}	0.52	0.67	0.50	0.61	0.61	0.50	0.59	0.62	0.40
	$F_{B^3}^*$	0.57	0.75	0.60	0.71	0.69	0.61	0.64	0.71	0.50
	P_{B^3}	0.39	0.77	0.73	0.61	0.79	0.59	0.48	0.72	0.90
	R_{B^3}	1.00	0.72	0.51	0.84	0.60	0.63	0.96	0.71	0.35

Results

Measure	Baseline	VIPS	HEPS	Corm.	MMD.	Meier	MV@1	MV@2	MV@3	MV@4
Segments	1.0	16.1	36.1	15.3	23.0	4.6	6.5	18.7	36.5	69.5
pixels	F_{B^3}	0.24	0.38	0.33	0.36	0.42	0.32	0.30	0.39	0.30
	$F_{B^3}^*$	0.28	0.47	0.44	0.53	0.54	0.50	0.35	0.50	0.45
	P_{B^3}	0.16	0.36	0.36	0.39	0.51	0.48	0.22	0.38	0.60
	R_{B^3}	1.00	0.67	0.56	0.80	0.57	0.52	0.96	0.72	0.36
chars	F_{B^3}	0.52	0.67	0.50	0.61	0.61	0.50	0.59	0.62	0.40
	$F_{B^3}^*$	0.57	0.75	0.60	0.71	0.69	0.61	0.64	0.71	0.50
	P_{B^3}	0.39	0.77	0.73	0.61	0.79	0.59	0.48	0.72	0.90
	R_{B^3}	1.00	0.72	0.51	0.84	0.60	0.63	0.96	0.71	0.35

Conclusion

- ❑ Empirical evaluation of
 - 5 web page segmentation algorithms on
 - 8490 web pages
- ❑ Usage of web archiving technology for reproducibility
- ❑ VIPS performs best overall, but not for *pixel* segments
- ❑ Competitive performance for purely visual approaches
- ❑ When fitted to DOM nodes, also a generic object detection algorithm trained on photos performs competitively