

# User Models, Metrics and Measures of Search: A Tutorial on the CWL Evaluation Framework

## ACM CHIIR UMMMS 2021

by

Leif Azzopardi, Alistair Moffat, Paul Thomas and Guido Zuccon



# **Section Three:**

## **C/W/L extensions and open research**

Presenter: Paul Thomas

# This section

Web search ain't like that: measuring a modern SERP

Metric evaluation: how do we know it's any good?

Directions: how far can we push this thing?

Microsoft Bing

new york time

Recent searches: new york times

ALLWORKIMAGESVIDEOSMAPSNEWSSHOPPING

359,000 ResultsDateLanguageRegion

New YorkTimeThings to doWeatherEventsItinerariesWhat to eatTravel guide

Microsoft

Who can see work results?

MSLibrary: News Hub

https://microsoft.sharepoint.com/sites/library/Sitep...

See all results

Send feedback

Current time in New York, United States (UTC-5)

6:26 AM

Tuesday, 9 March 2021

**The New York Times - Breaking News, US News, World News ...**  
<https://www.nytimes.com>  
Live news, investigations, opinion, photos and video by the journalists of **The New York Times** from more than 150 countries around the world. Subscribe for coverage of U.S. and international news ...

#### Crossword

Play the Daily New York Times Crossword puzzle edited by Will Shortz online. Try free ...

#### World

A court in Colombia is exposing atrocities in the country's long civil war, trying to compile ...

#### Opinion

New York Times Opinion columnists, editorials and op-eds. Analysis from David Brooks, ...

#### Politics

By The New York Times Here's How the Senate Pared Back Biden's Stimulus Plan The \$1.9 ...

#### Business

Exclusive reporting and breaking news about the most important business and ...

#### Sports

Find breaking news & sports news on the NFL, the NBA, the NCAA, the NHL, baseball, golf, ...

#### U.S.

View the latest news, top stories, headlines, photos and videos from around the nation. ...

#### Books

For the first time in more than a century, the society is adding new spots for members, ...

#### Style

Photo Credit Illustration by Shannon Lin/The New York Times The Awkward but Essential ...

NEWS FROM NYTIMES.COM

Pandemic Relief Bill Fulfills Biden's Promise to ...

1 hour ago

Cornel West Is Leaving Harvard After Tenure ...

13 hours ago

Prince Harry Finally Takes On White Privilege: His Own

20 hours ago

#### Current Local Time in New York, New York, USA

<https://www.timeanddate.com/worldclock/usa/new-york>  
8/6/2020 · Current local time in USA – New York – New York. Get New York's weather and area codes, time zone and DST. Explore New York's sunrise and sunset, moonrise and moonset.

#### Current local time in New York, United States

[https://www.worldtimeserver.com/current\\_time\\_in\\_US-NY.aspx](https://www.worldtimeserver.com/current_time_in_US-NY.aspx)

### New York City

City in New York State

New York City, often called simply New York, is the most populous city in the United States. With an estimated 2019 population of 8,336,817 distributed over about 302.6 square miles, New York City is also the most densely populated major city in the United States. Located at the southern tip of the State of New York, the city is the center of the New York met...

W

Wikipedia

Official site

Local time: 6:26 am 9/3/2021

Population: 8.399 million (2018)

Area: 1,213 km²

Travel tip: The first time you go to New York, go ahead and be a sight

Mayor: Bill de Blasio

Nearby airports: Newark Liberty International Airport · LaGuardia Airport · John F. Kennedy International Airport

#### Plan a trip

Travel guide

Explore New York

Book flights

Search by price, schedule, and airline

Book hotels

Avg daily rate for 3-star hotel: \$218

4-day itinerary

Statue of Liberty · One World Trade Center

#### Demographics

Year	Pop	±%
1698	4,937	—
1712	5,840	+18.3%
1723	7,248	+24.1%
See more		

#### People also search for

See all (20+)

puzzle edited by Will Shortz online. Try free ...

#### World

A court in Colombia is exposing atrocities in the country's long civil war, trying to compile ...

#### Opinion

New York Times Opinion columnists, editorials and op-eds. Analysis from David Brooks, ...

#### Politics

By The New York Times Here's How the Senate Pared Back Biden's Stimulus Plan The \$1.9 ...

#### Business

Exclusive reporting and breaking news about the most important business and ...

Search

NEWS FROM NYTIMES.COM

Pandemic Relief Bill Fulfills Biden's Promise to ...

1 hour ago

Cornel West Is Leaving Harvard After Tenure ...

13 hours ago

Prince Harry Finally Takes On White Privilege: His Own

20 hours ago

#### Current Local Time in New York, New York, USA

<https://www.timeanddate.com/worldclock/usa/new-york>  
8/6/2020 · Current local time in USA – New York – New York. Get New York's weather and area codes, time zone and DST. Explore New York's sunrise and sunset, moonrise and moonset.

#### Current local time in New York, United States

[https://www.worldtimeserver.com/current\\_time\\_in\\_US-NY.aspx](https://www.worldtimeserver.com/current_time_in_US-NY.aspx)  
Try our World Meeting Planner and get a color-coded chart comparing the time of day in New York, United States with all of the other international locations where others will be participating. Time Zone News . New York to set clocks back on November 1, 2015; New York to set clocks forward on March 13 ; Put a clock on your blog! New York, United States . Have your own blog or web site? Put a ...

#### Coronavirus in the U.S.: Latest Map ... - The New York Times

<https://www.nytimes.com/interactive/2020/us/coronavirus-us-cases.html>  
20/2/2021 · The New York Times is engaged in a comprehensive effort to track information on every coronavirus case in the United States, collecting information from ...

#### The New York Times in Print for Wednesday, Jan. 20, 2021 ...

<https://www.nytimes.com/issue/todayspaper/2021/01/20/todays-new-york-times>  
20/1/2021 · Photo Credit Jason Andrew for The New York Times 'This Kettle Is Set to Boil': New Evidence Points to Riot Conspiracy. While most arrests in the Capitol riot have been individuals, new charges ...

#### Covid-19 Vaccine Tracker Updates: The ... - The New ...

8 mins read  
<https://www.nytimes.com/interactive/2020/science/coronavirus-vaccine...>  
5/3/2021 · The Trump administration awarded a \$1.9 billion contract in July for 100 million doses, but The New York Times reported in December that the administration passed ...

#### Tiles - The New York Times - The New York Times

<https://www.nytimes.com/puzzles/tiles>  
About New York Times Games. Times games have captivated solvers since the launch of the Crossword in 1942. Our experts create engaging word and visual games — in 2014 we introduced the Mini ...

#### Related searches for new york time

- time now in new york city
- new york times news
- new york time zone
- new york state local time
- current time in new york
- current time new york city

the NBA, the NCAA, the NHL, baseball, golf, ...

#### U.S.

View the latest news, top stories, headlines, photos and videos from around the nation. ...

#### Books

For the first time in more than a century, the society is adding new spots for members, ...

#### Style

Photo Credit Illustration by Shannon Lin/The New York Times The Awkward but Essential ...

W

Wikipedia Official site

Local time: 6:26 am 9/3/2021

Population: 8.399 million (2018)

Area: 1,213 km²

Travel tip: The first time you go to New York, go ahead and be a sight

Mayor: Bill de Blasio

Nearby airports: Newark Liberty International Airport · LaGuardia Airport · John F. Kennedy International Airport

Plan a trip

Travel guide

Explore New York

Book flights

Search by price, schedule, and airline

Book hotels

Avg daily rate for 3-star hotel: \$218

4-day itinerary

Statue of Liberty · One World Trade Center

Demographics

Year	Pop	±%
1698	4,937	—
1712	5,840	+18.3%
1723	7,248	+24.1%
See more		

People also search for

See all (20+)

Los Angeles

Chicago

London

Washington, D.C.

Boston

Explore more

Largest cities by population in the United States

Data from: Wikipedia · Freebase · Kid101 · TripHobo · Skyscanner · TripAdvisor · Booking.com

Wikipedia text under CC-BY-SA license

Suggest an edit



**WEB SEARCH AIN'T LIKE THAT**

# Different results

 **Current Local Time in New York, New York, USA**

Time/General

Weather ▾

Time Zone

DST Changes

Sun & Moon ▾



**5:18:27 am EST**  
Sunday, 7 March 2021  
[Fullscreen](#)

**Country:** [United States](#)

**State:** [New York \(NY\)](#)

**Lat/Long:** [40°43'N / 74°01'W](#)

**Elevation:** 8m

**Currency:** [United States Dollar \(USD\)](#)

**Languages:** [English](#)

**Dial Codes:** [+1 - See how to dial](#)



 Sunday, 14 March 2021 — Daylight Saving Time Starts. Clocks move forward 1 hour. [See more](#)

**Weather**  
-2 °C  
Overcast.  
4 / -3 °C

  
Mon 8.  5 / -4 °C

  
**Time Zone**

  
**DST starts**

  
**DST ends**

  
**Difference**

## Eastern Time Zone

From Wikipedia, the free encyclopedia

*This article is about Eastern Standard Time (UTC−5) and Eastern Daylight Time in the Americas. For Eastern Standard Time in Australia (UTC+10), see [Australian Eastern Standard Time](#). For the Eastern Daylight Time Zone of Australia, see [Daylight saving time in Australia](#).*

The **Eastern Time Zone** (ET) is a [time zone](#) encompassing part or all of 23 [states](#) in the eastern part of the [United States](#), parts of eastern Canada, the state of [Quintana Roo](#) in Mexico, Panama and Colombia, mainland Ecuador, Peru, and a small portion of westernmost Brazil in South America, along with certain [Caribbean](#) and Atlantic islands.

Places that use:

- **Eastern Standard Time (EST)**, when observing [standard time](#) (autumn/winter), are five hours behind [Coordinated Universal Time](#) (UTC−05:00).
- **Eastern Daylight Time (EDT)**, when observing [daylight saving time](#) (spring/summer), are four hours behind Coordinated Universal Time (UTC−04:00).

In the northern parts of the time zone, on the second Sunday in March, at 2:00 a.m. EST, clocks are advanced to 3:00 a.m. EDT leaving a one-hour "gap". On the first Sunday in November, at 2:00 a.m. EDT, clocks are moved back to 1:00 a.m. EST, thus "duplicating" one hour. Southern parts of the zone (Panama and the Caribbean) do not observe daylight saving time.

The largest city in the Eastern Time Zone is [New York](#); the [New York metropolitan area](#) is the largest metropolitan area in the zone.

### Contents [\[hide\]](#)

- [History](#)
- [Daylight saving time](#)
- [Canada](#)
- [United States](#)
- [Mexico](#)
- [Caribbean islands](#)
  - [Turks and Caicos Islands](#)
- [South America](#)
- [See also](#)
- [References](#)
- [External links](#)

[History](#) [\[edit\]](#)

**Eastern Time Zone**

time zone



EST  
UTC−05:00

EDT  
UTC−04:00

**Current time**

04:43, 7 March 2021 [EST](#) [\[refresh\]](#)

**Observance of DST**

DST is observed in parts of this time zone.

# Gain from snippets

## Time in New York

City

The state of New York uses the Eastern Time Zone with daylight saving time. Because the largest city in the state, New York City, is the largest city in the United States, and because the Eastern Time Zone is used in the nation's capital, Washington, D.C., in addition to the fact that a majority of the nation's population is in that time zone, Eastern Time is the most ... [+](#)



Wikipedia

**People also search for:** [Time in Arizona](#) · [Time in Indiana](#) · [Time in Oregon](#) · [Time in Idaho](#) · [Time in Florida](#) · [Time in Nevada](#) · [Time in Kansas](#) [+](#)

Data from: Wikipedia

Wikipedia text under CC-BY-SA license

[Suggest an edit](#)

# Heterogeneity

Related searches for new york time

**nyse opening hours**

**world clock**

time in **san francisco**

time in **los angeles now**

**what time is it in us**

**sydney vs** new york time

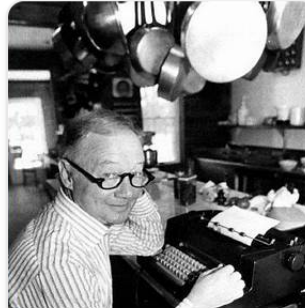
time in new york **sydney**

time **now** in new york **city**

The New York Times columnists



Thomas  
Friedman



Craig Claiborne



Robert H. Frank

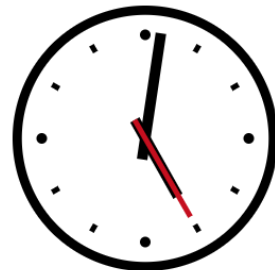


Lisa Sanders



Tyler Cowen

Current time in New York, United States (UTC-5)



5:01 AM

Saturday, 6 March 2021

# METRIC EVALUATION

“

It is axiomatic that a good metric is one that has a high correlation with user satisfaction ratings. ”

# Metric evaluation: Cost

Labelling takes real time and money.

Computation may also take real time and money.



# Metric evaluation: Sensitivity



Metrics are more useful the more sensitive they are to real changes.



# Metric evaluation: Interpretability

Interpretability, debuggability, use for diagnostics are extremely important.

Can we **do something** with the result?



# Metric evaluation: Face validity



Does the metric look like it should measure what we want it to measure?

Does it look effective?

Do we believe the user model?

# Metric evaluation: External validity

How well do our conclusions carry over to another setting?

How much do we learn about the “real world”, not just the instrument itself?





# Metric evaluation: External validity

Correspondence with...

- Other metrics
- Short-term behaviours
- Side-by-side/setwise pref's
- Long-term behaviours
- Business goals



# Reflecting observable behaviour: 1

Stopping

Accumulated utility

Accumulated cost

Likelihood of stopping at  $i$   
 $= L_i$

Model / measure	Mean likelihood
P@1	0.49
P@5	0.00
P@10	0.00
SDCG@1	0.49
SDCG@5	0.33
SDCG@10	0.33
RR	0.36
RBP, $\phi=0.1$	0.44
RBP, $\phi=0.7$	0.16
INST, T=1	0.34
INST, T=2	0.21
IFT	0.76

---

# Reflecting observable behaviour: 2

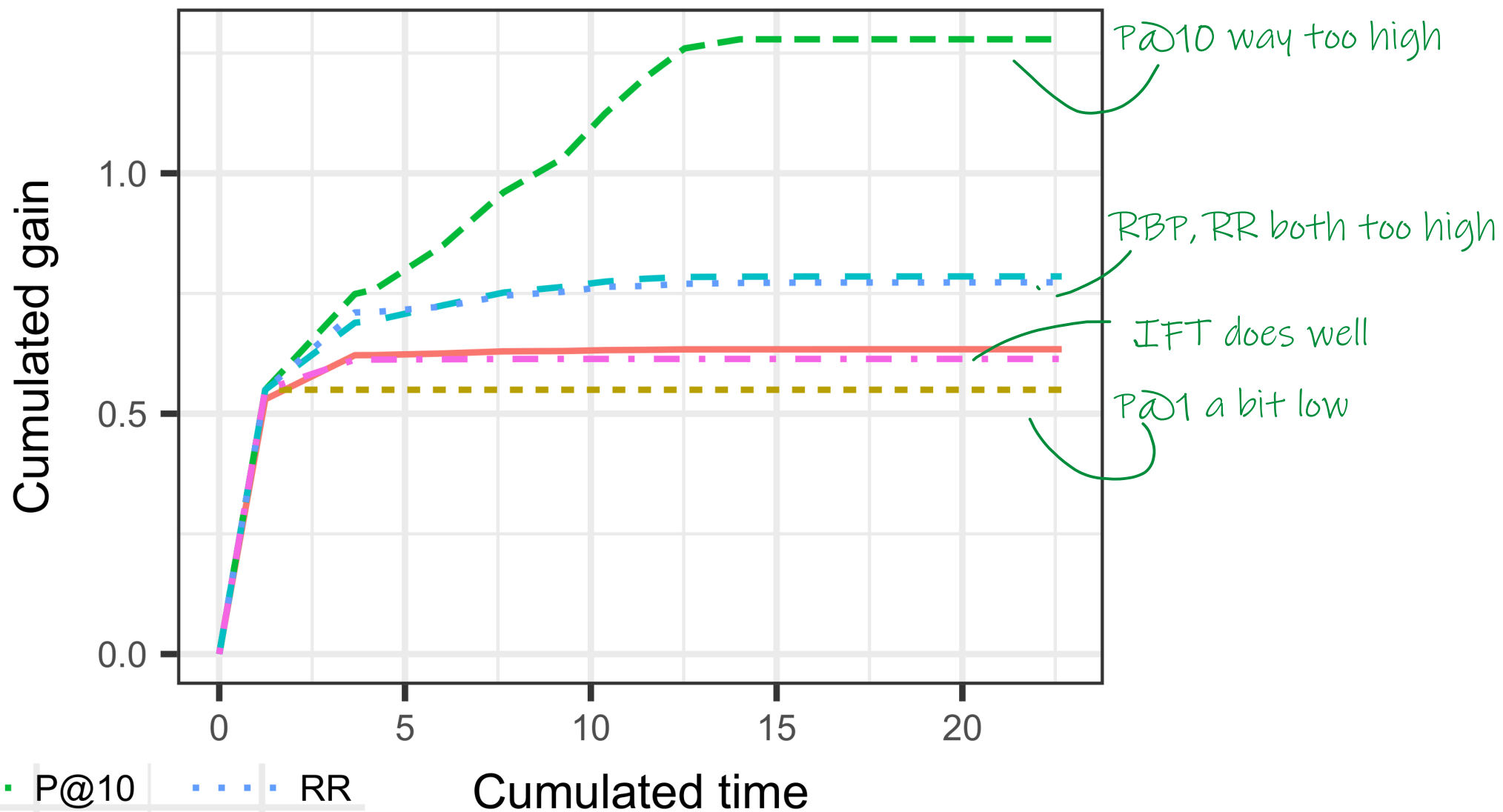
Stopping

Accumulated utility

Accumulated cost

Expected total utility

$$= \sum_i \left( L_i \sum_{j=1}^i r_j \right)$$



Average over 5000 impressions  
Azzopardi et al., SIGIR'18



# Reflecting observable behaviour: 3

Stopping

Accumulated utility

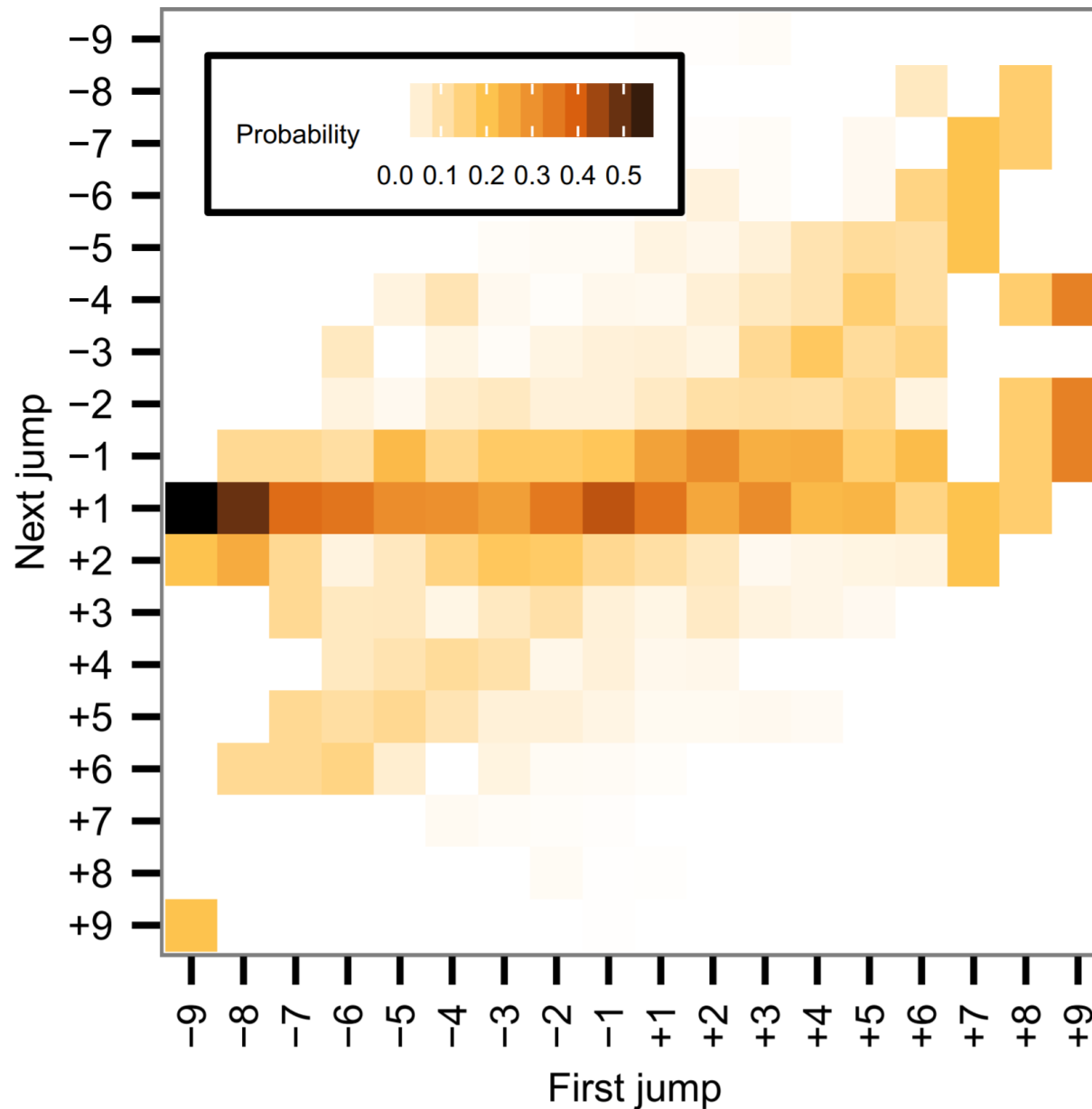
Accumulated cost

Expected total cost

$$= \sum_i \left( L_i \sum_{j=1}^i k_i \right)$$

## **DIRECTIONS**

# Top-down reading?



# Top-down reading??


Microsoft Bing

tom cruise

Recent searches: chair 2021

ALLWORKIMAGESVIDEOSMAPSNEWSSHOPPING

1,100,000 ResultsDate▼Language▼Region▼

Tom Cruise

MoviesUpcoming moviesAwardsQuotesRecent highlightsInterviewsFashionMerchandiseBiography

Tom Cruise - Wikipedia

https://en.wikipedia.org › wiki › Tom\_Cruise

Overview


Early life

Career

Personal life

Legacy

Litigati




Thomas Cruise Mapother IV (born July 3, 1962) is an American actor and producer. He has received various accolades for his work, including three Golden Globe Awards and three nominations for Academy Awards. He is one of the highest-paid actors in the world. His films have grossed over \$4 billion in North America and over \$10.1 billion worldwide, making him one of the highest-grossing box office stars of all time.

Wikipedia · Text under CC-BY-SA license


Tom Cruise - IMDb

https://www.imdb.com › name › nm0000129


In 1976, if you had told fourteen-year-old Franciscan seminary student **Thomas Cruise Mapother IV** that one day in the not too distant future he would be **Tom Cruise**, one of the top 100 movie...



The IMDb Show – This Summer, Tom Cruise is back as super-spy Ethan ...



The IMDb Show – From the red carpet of their Paris premiere, Tom Cruise, ...



IMDbrief – On this IMDbrief, we break down how director Leigh ...

Videos of Tom Cruise

bing.com › videos


80s

Movies List

High School


Full Movies

Old




Tom Cruise's COVID-19 Rant Causes 5 Crew Members To Quit

412K views · 1 month ago  
YouTube · ET Canada



Tom Cruise Lashes Out at Mission: Impossible 7 Crew After They Break C...

859K views · 1 month ago  
YouTube · Inside Edition



Tom Cruise Lashes Out at Mission: Impossible 7 Crew After They Break C...

106K views · 1 month ago  
YouTube · Entertainment...

See more videos

Official Tom Cruise Website

www.tomcruise.com

website

email

site

news

TOM CRUISE is a global cultural icon who has made an immeasurable impact on cinema by creating some of the most memorable characters of all time.

See Microsoft results for tom cruise

Tom Cruise - Biography - IMDb

https://www.imdb.com › name › nm0000129 › bio

Tom is the only son (among four children) of nomadic parents, Mary Lee (Pfeiffer), a special

Tom Cruise

American Actor

Thomas Cruise Mapother IV, better known by the stage name Tom Cruise, is an American actor and producer. He has received various accolades for his work, including three Golden Globe Awards and three nominations for Academy Awards. He is one of the highest-paid actors in the world. His films have grossed over \$4 billion in Nort...

Wikipedia

IMDb

Instagram

Facebook

YouTube

Born: 3 Jul 1962 (age 58) · Syracuse, NY

Height: 170 cm


Net worth: 600 million USD (2020)

Spouse: Katie Holmes (m. 2006 - 2012) · Nicole Kidman (m. 1990 - 2001) · Mimi Rogers (m. 1987 - 1990)


Partner: Penélope Cruz (2001 - 2004)

Upcoming movies: Top Gun: Maverick · Mission: Impossible 7 · Mission: Impossible 8


Recent Highlights



Tom Cruise slams 'Mission: Impossible 7' 21/12/2020 · Variety and 16 more



Tom Cruise visits London movie theater to watch 'Tenet,' and the video of his trip 27/8/2020 · MSN and 18 more




actor Doug Gray Scott says Tom Cruise made him give up Wolverine role, saying he couldn't 10/3/2020 · MSN and 2 more


See more

Movies


See all (20+)




Top Gun: Maverick 2021




Mission: Impossible - Fallout 2018



Top Gun 1986




Edge of Tomorrow 2014




Mission: Impossible 7 2021

Romance


See all (5+)




Romance movie 1




Romance movie 2



Romance movie 3

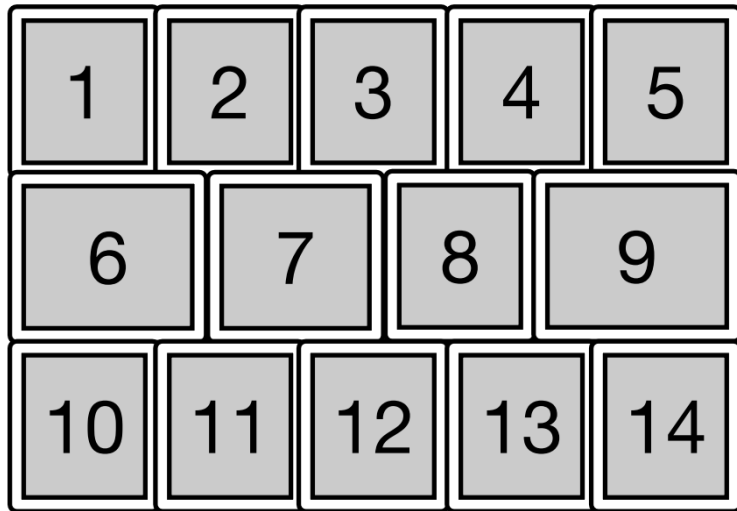


Romance movie 4



Romance movie 5

# Top-down reading???

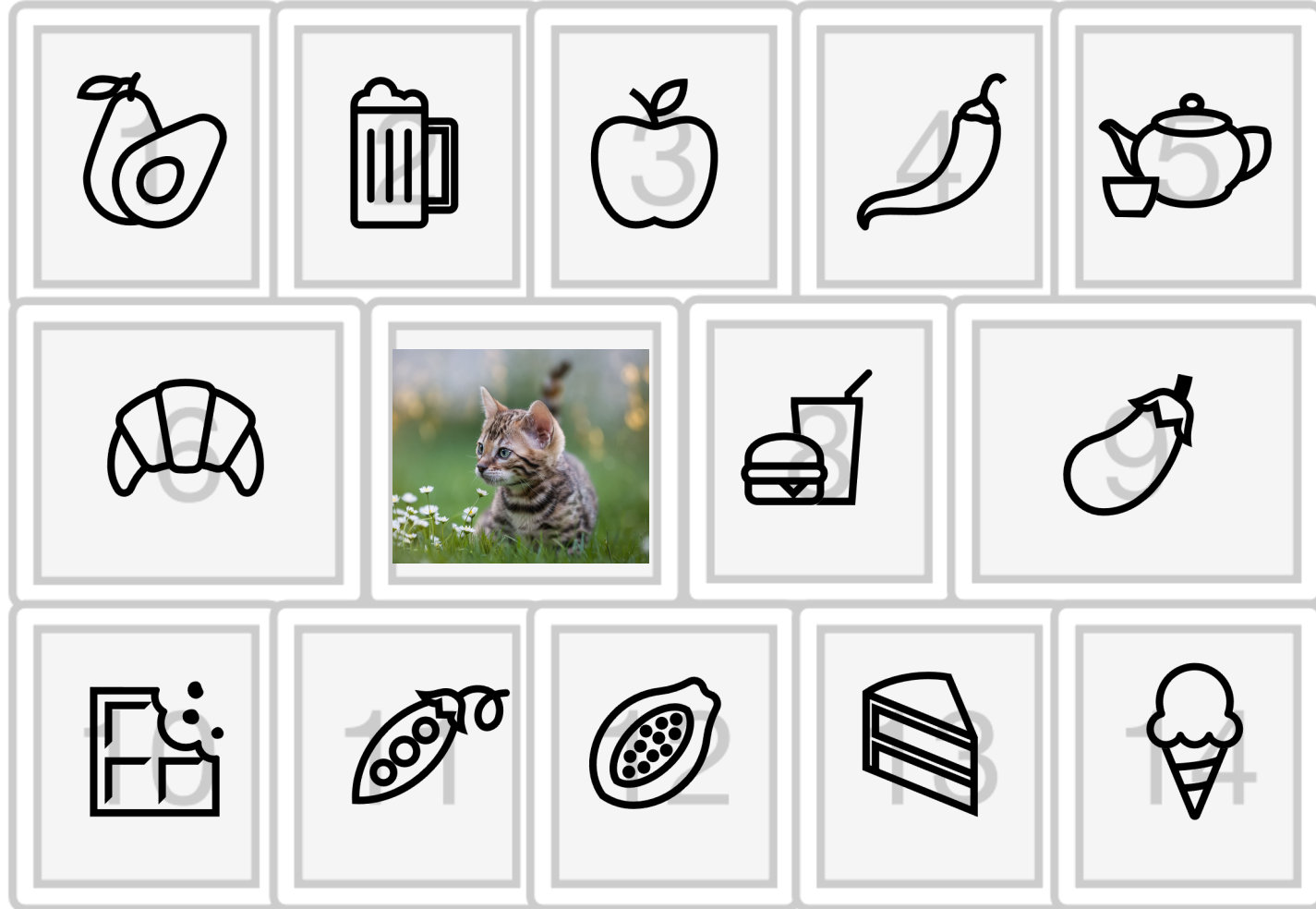


‘Z’: (1,2,3,4,5, 6,7,8,9, 10,11,12,13,14)

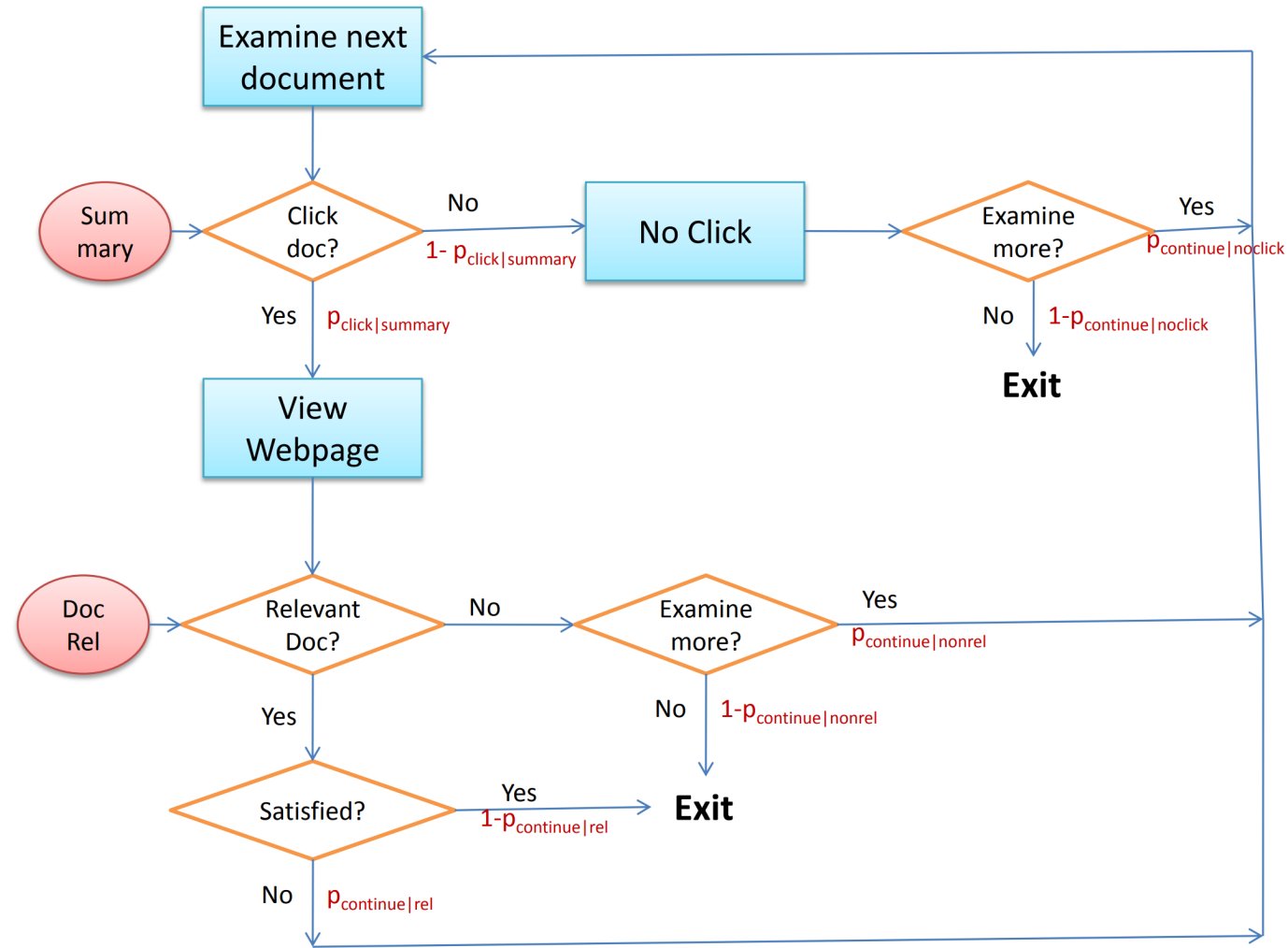
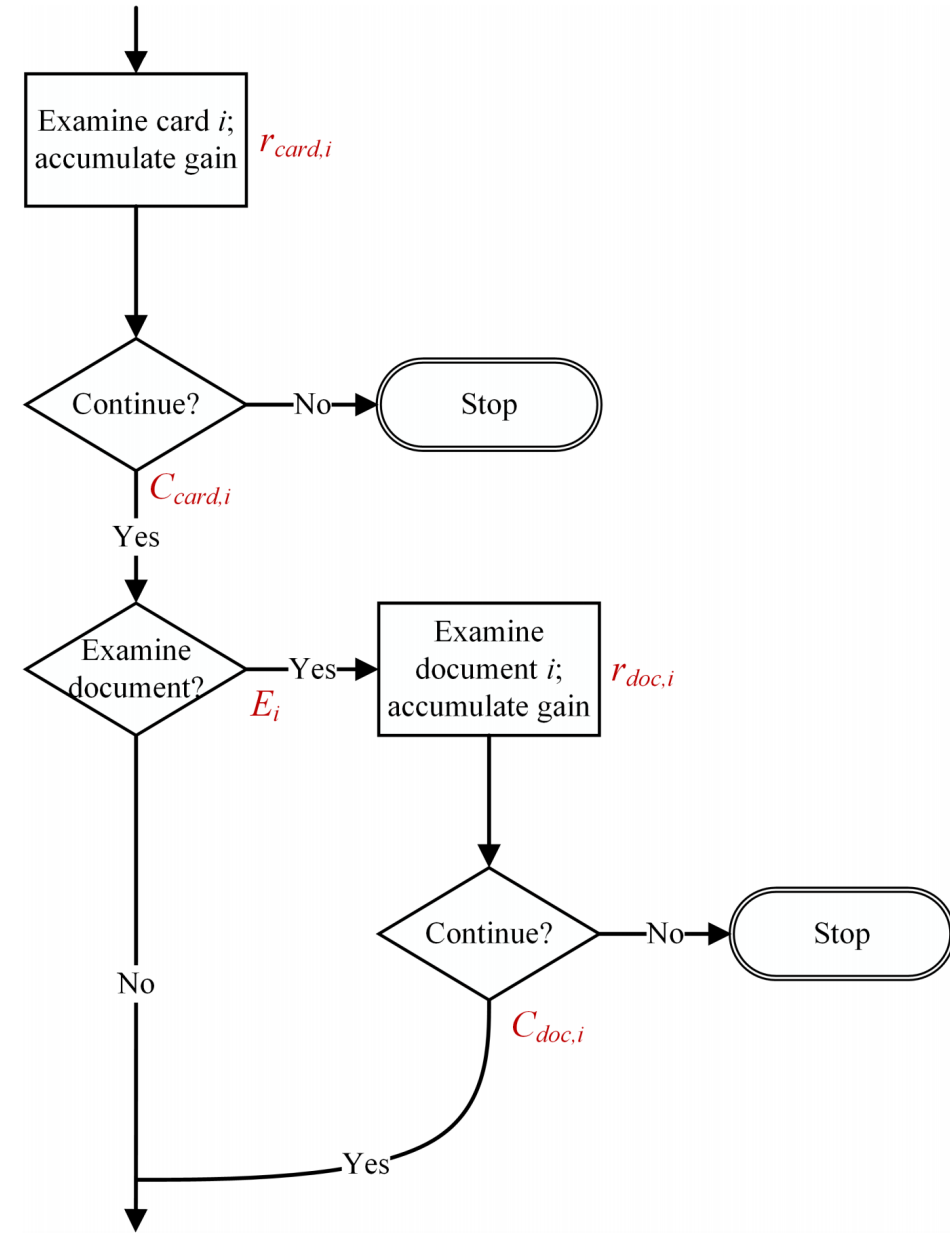
‘S’: (1,2,3,4,5, 9,8,7,6, 10,11,12,13,14)

‘T’: (3,2,4,1,5, 7,8,6,9, 12,11,13,10,14)

# Top-down reading???



# Gain from cards

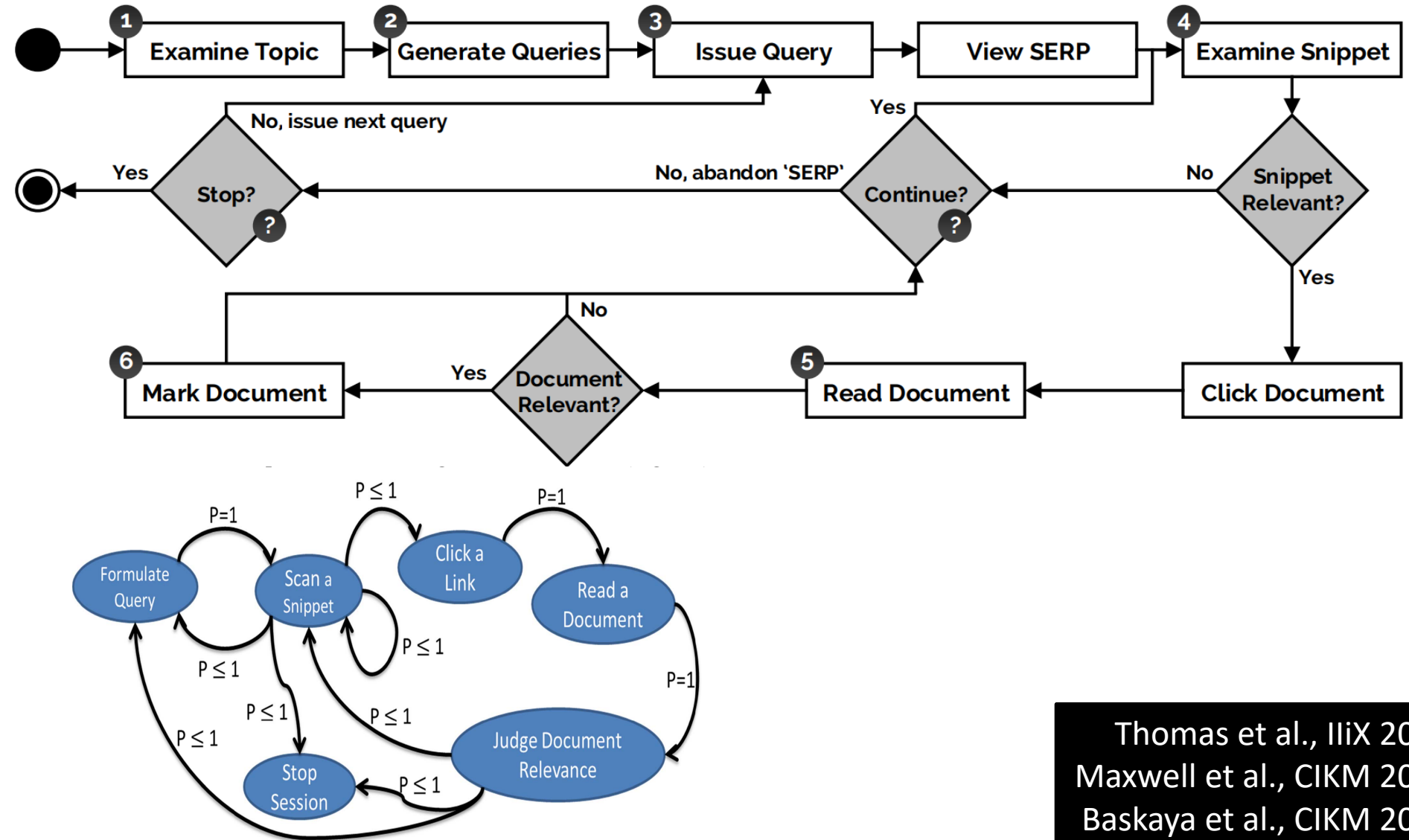
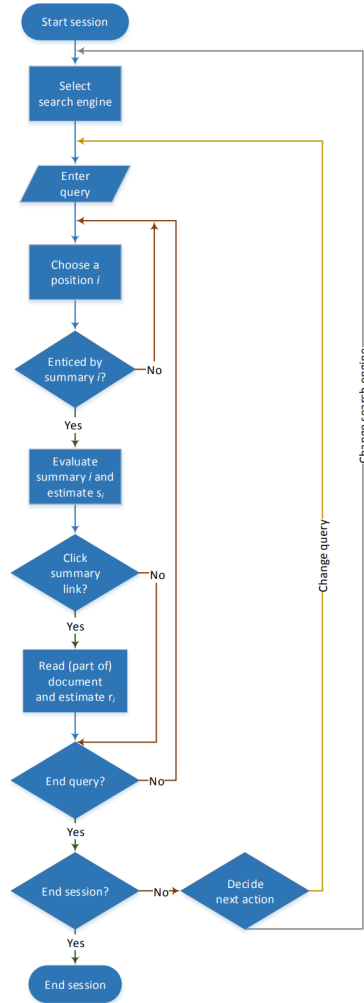


# Heterogeneity

Type	Position	
	Core	Right Rail
Intercept		3.65
Algorithmic web results	1.00	-
Advertisements	1.49	0.30
News	5.62	-
Query suggest.	1.41	-
Images	0.96	-
Videos	3.91	-
Entity cards	8.91	0.45
Result disambig.	-	1.81
Stock	0.97	-
Other	3.22	0.96



# Sessions



Thomas et al., IliX 2014  
 Maxwell et al., CIKM 2015  
 Baskaya et al., CIKM 2013

# Other directions



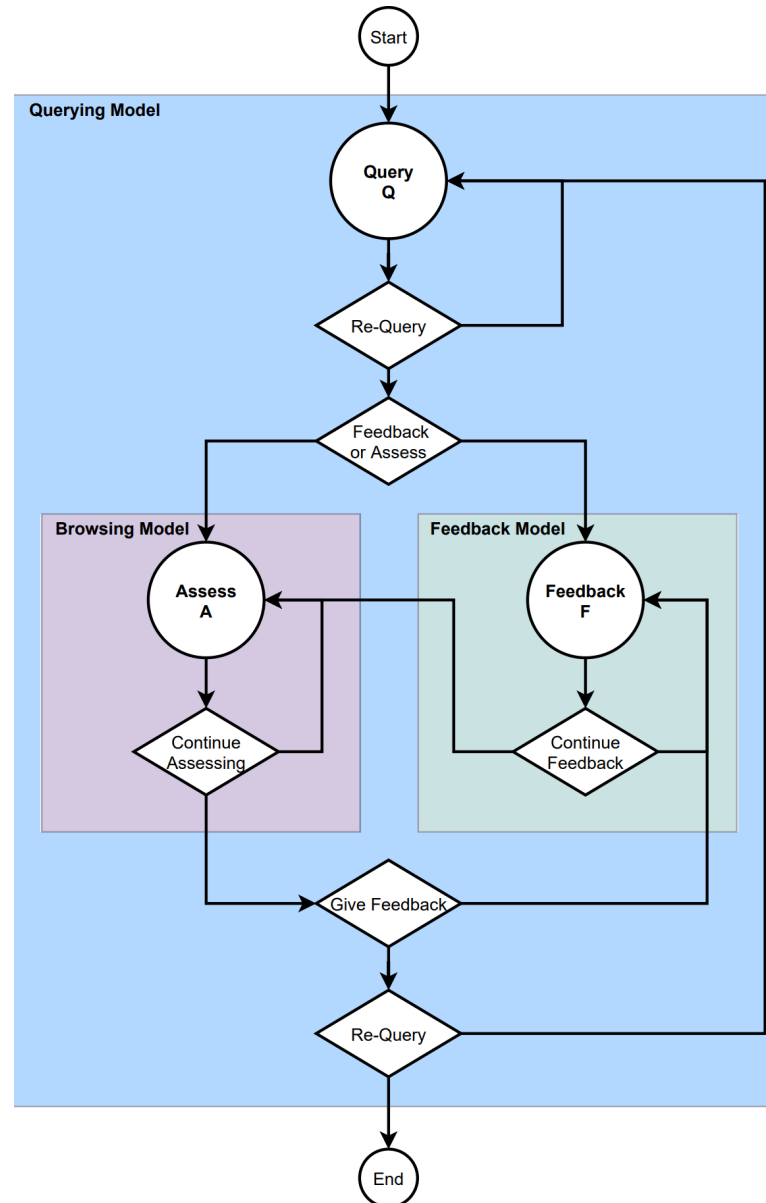
Conversational interactions

CWL during an interaction

Streaming documents

Recommendation

# Other directions



# Other directions



Conversational interactions

CWL during an interaction

Streaming documents

Recommendation

# Summary of section 3

Web search ain't like that: measuring a modern SERP

Metric evaluation: how do we know it's any good?

Directions: how far can we push this thing?