

Isaac Johnson -- Research Scientist -- Wikimedia Foundation

WikiConference North America -- 8 October 2021

Etherpad: https://etherpad.wikimedia.org/p/wcna2021-4202

The Research Team



<u>Leila Zia</u> Director, Head of Research



<u>Pablo Aragón</u> Research Scientist



Martin Gerlach Research Scientist



<u>Isaac Johnson</u> Research Scientist



<u>Fabian Kaelin</u> Senior Research Engineer



Emily Lescak Senior Research Community Officer



Miriam Redi Senior Research Scientist



Diego Sáez-Trumper Senior Research Scientist

...and many formal collaborators: https://w.wiki/4BNx

Learn more: https://research.wikimedia.org

Modern art (Q38166)

Ż_A 50 languages ∨

Artistic works produced during the period extending roughly from the 1860s to the 1970s. From Wikipedia, the free encyclopedia

For other uses, see Modern art (disambiguation). Not to be confused with contemporary art, nor art moderne.

Modern art includes artistic work produced during the period extending roughly from the 1860s to the 1970s, and denotes the styles and philosophies of the art produced during that era. [1] The term is usually associated with art in which the traditions of the past have been thrown aside in a spirit of experimentation. [2] Modern artists experimented with new ways of seeing and with fresh ideas about the nature of materials and functions of art. A tendency away from the narrative, which was characteristic for the traditional arts, toward abstraction is characteristic of much modern art. More recent artistic production is often called contemporary art or postmodern art.

Modern art begins with the heritage of painters like Vincent van Gogh, Paul Cézanne, Paul Gauguin, Georges Seurat and Henri de Toulouse-Lautrec all of whom were essential for the development of modern art. At the beginning of the 20th century Henri Matisse and several other



Agenda

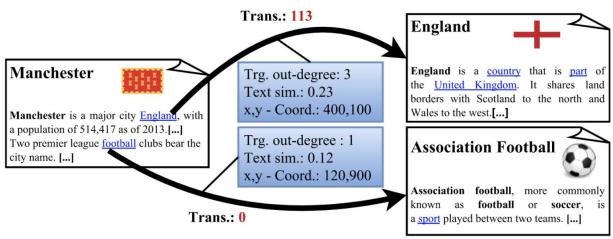
- What makes a successful link?
- 2. What is the relationship between links and gender equity?
- 3. What is the relationship between links and machine learning coverage?
- 4. What is the relationship between links and **new editors**?

Link Success: Research

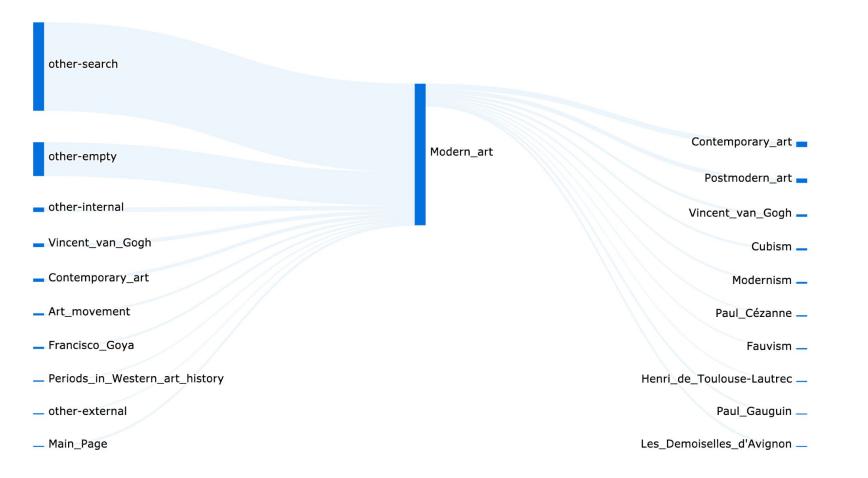
What Makes a Link Successful on Wikipedia?

Dimitar Dimitrov, Philipp Singer, Florian Lemmerich, Markus Strohmaier. WWW 2017. [arXiv]

- About 4% of all links on English Wikipedia are clicked 10+ times per month
- Page placement: top + upper-left (for English as LTR language)
- Centrality: more peripheral articles
- Semantics: topically-similar articles

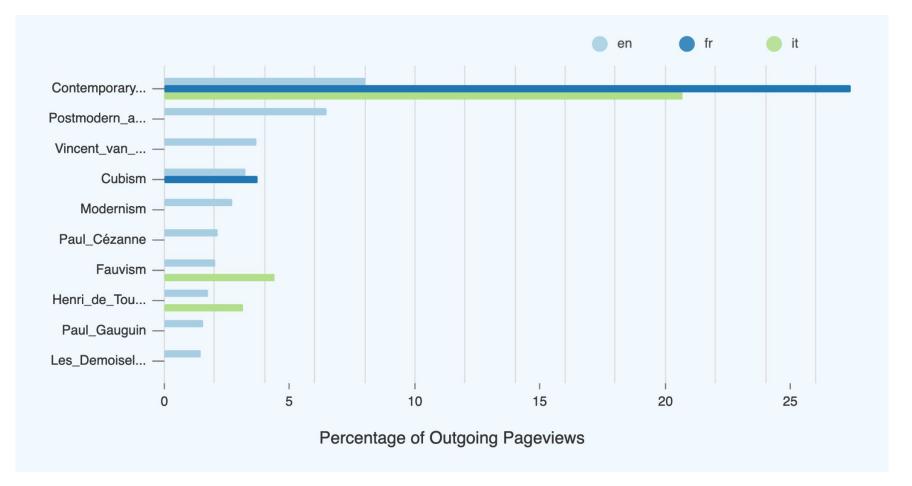


Link Success: Clickstream Data



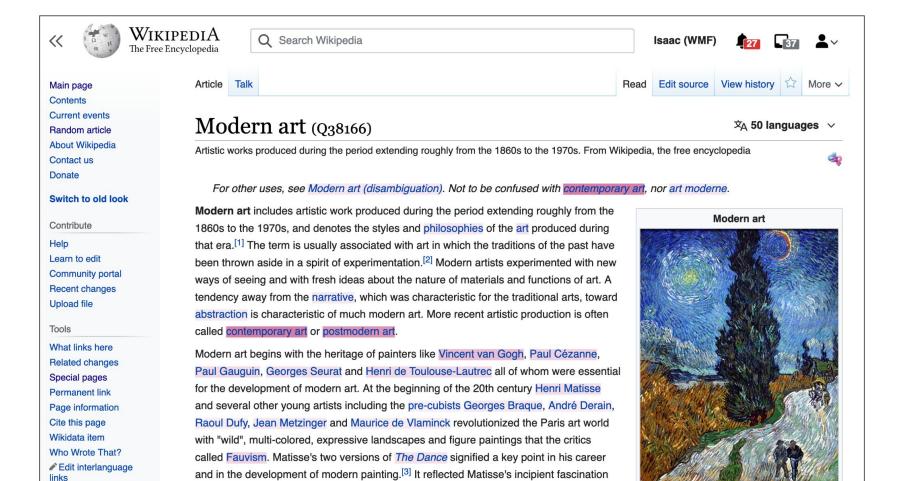
Explore further: https://wikinav.toolforge.org/

Details: https://techblog.wikimedia.org/2021/09/17/analyzing-the-wikipedia-clickstream-just-got-easier-with-wikinav/



Explore further: https://wikinav.toolforge.org/

Details: https://techblog.wikimedia.org/2021/09/17/analyzing-the-wikipedia-clickstream-just-got-easier-with-wikinav/



with primitive art: the intense warm color of the figures against the cool blue-green

Vincent van Gogh, Country Road in Provence

Gender Equity + Links: Research

Women through the glass ceiling: gender asymmetries in Wikipedia

Claudia Wagner, Eduardo Graells-Garrido, David Garcia & Filippo Menczer. EPJ Data Science 2016. [pdf]

- Examined the gender gap through different lenses: content policies, linguistic bias, and structural properties
- Women have lower PageRank values than men

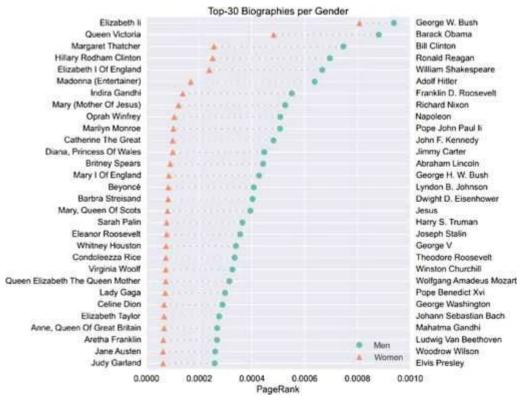
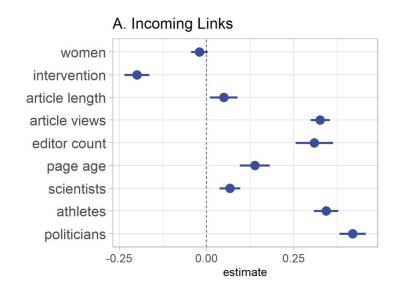


Figure 5. Top 30 biographies sorted by PageRank.

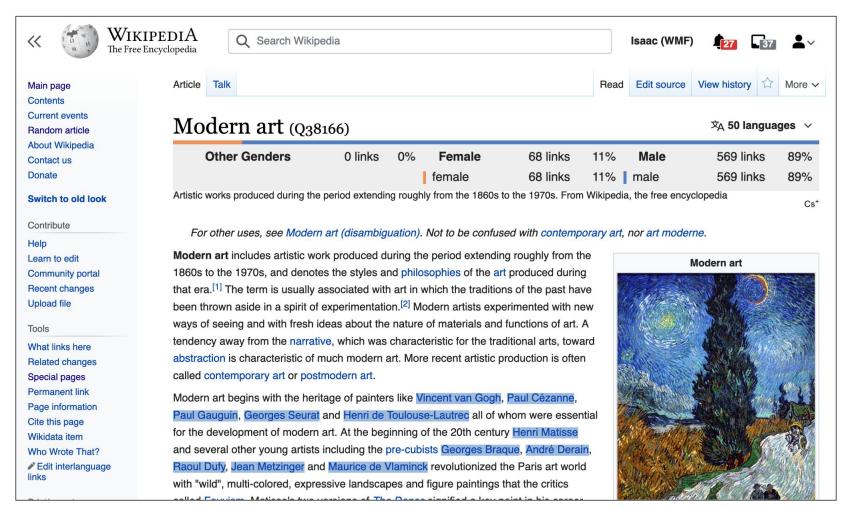
The Gender Divide in Wikipedia: Quantifying and assessing the impact of two feminist interventions

Isabelle Langrock and Sandra González-Bailón. SSRN 2021. [pdf]

- Art+Feminism and 500 Women
 Scientists have been very successful at creating content about women on English Wikipedia.
- Less attention has been directed at linking this new content in with the rest of Wikipedia.



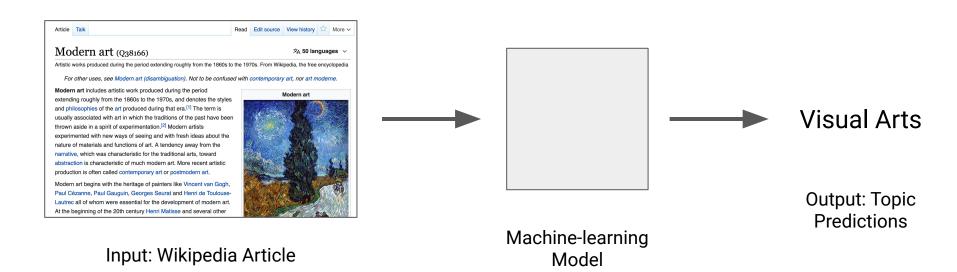
Gender Equity + Links: Data



Links + Machine Learning Coverage

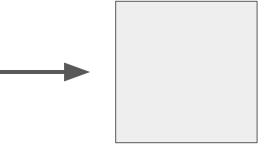
Language-agnostic Topic Classification for Wikipedia

Isaac Johnson, Martin Gerlach, Diego Sáez-Trumper. WikiWorkshop 2021. [arXiv]





Input: **English** Wikipedia Article



Visual Arts

Output: Topic Predictions

Machine-learning Model Bài viết Thảo luận Đọc Sửa đổi Sửa mã nguồn ☆ Thêm ✔ Tìm kiếm Wikipedia Q

đầu thế lợ 00. Hong Motioco và một cố nghệ cĩ trả khác học gầm các nghệ cĩ tiền lập thể như Coorgo

Nghệ thuật hiện đại (Q38166)

Bách khoa toàn thư mở Wikipedia

Nghệ thuật hiện đại (tiếng Anh: modern art) bao hàm các tác phẩm nghệ thuật được sáng tác trong giai đoạn khoảng từ thập niên 1860 đến thập niên 1970, biểu thị các phong cách và triết lý của nghệ thuật thời kỳ đó.[1] Nghệ thuật hiện đại gạt sang một bên các truyền thống quá khứ với một tinh thần thể nghiệm.[2] Các nghệ sĩ hiện đại thể nghiệm với những cách nhìn mới và với những ý tưởng mới về bản chất vật liệu và các chức năng của nghệ thuật. Nghệ thuật hiện đại có xu hướng xa rời tính tự sự (vốn dĩ là đặc trưng của nền nghệ thuật truyền thống) để hướng tới tính trừu tượng. Tác phẩm nghệ thuật ra đời trong các năm gần đây thường được gọi là nghệ thuật đương đại hoặc nghệ thuật hậu hiện đại.

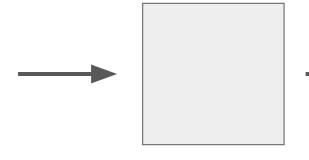
Nghệ thuật hiện đại bắt đầu với di sản hội họa của nhiều họa sĩ như Vincent van Gogh, Paul Cézanne, Paul Gauguin, Georges Seurat và Henri de Toulouse-Lautrec, tất cả mang tầm vóc quan trọng ảnh hưởng đến sự phát triển của nghệ thuật hiện đại. Vào



Bức *Tiếng thét* của Edvard Munch, 1893



Input: **Vietnamese** Wikipedia Article



_

???

Machine-learning Model

Output: Topic Predictions

Modern art (Q38166)

Ż_A 50 languages ∨

Artistic works produced during the period extending roughly from the 1860s to the 1970s. From Wikipedia, the free encyclopedia

For other uses, see Modern art (disambiguation). Not to be confused with contemporary art, nor art moderne.

Modern art includes artistic work produced during the period extending roughly from the 1860s to the 1970s, and denotes the styles and philosophies of the art produced during that era. [1] The term is usually associated with art in which the traditions of the past have been thrown aside in a spirit of experimentation. [2] Modern artists experimented with new ways of seeing and with fresh ideas about the nature of materials and functions of art. A tendency away from the narrative, which was characteristic for the traditional arts, toward abstraction is characteristic of much modern art. More recent artistic production is often called contemporary art or postmodern art.

Modern art begins with the heritage of painters like Vincent van Gogh, Paul Cézanne, Paul Gauguin, Georges Seurat and Henri de Toulouse-Lautrec all of whom were essential for the development of modern art. At the beginning of the 20th century Henri Matisse and several other



Modern art (Q38166)

文_A 50 languages ∨

Artistic works produced during the period extending roughly from the 1860s to the 1970s. From Wikipedia, the free encyclopedia

For other uses, see Q1205723. Not to be confused with Q186030, nor Q928960.

Modern art includes artistic work produced during the period extending roughly from the 1860s to the 1970s, and denotes the styles and Q1743282 of the Q735 produced during that era.^[1] The term is usually associated with art in which the traditions of the past have been thrown aside in a spirit of experimentation.^[2] Modern artists experimented with new ways of seeing and with fresh ideas about the nature of materials and functions of art. A tendency away from the Q1318295, which was characteristic for the traditional arts, toward Q673661 is characteristic of much modern art. More recent artistic production is often called Q186030 or Q3318563.

Modern art begins with the heritage of painters like Q5582, Q35548, Q37693, Q34013 and Q82445 all of whom were essential for the development of modern art. At the beginning of the 20th century Q5589 and several other young artists including the Q3408836



Đọc

Sửa đổi

Sửa mã nguồn

Thêm ~

Tìm kiếm Wikiped

Modern art (Q38166)

Artistic works produced during the period ext

For other uses, see Q1205723. Not t

Modern art includes artistic work produc extending roughly from the 1860s to the and Q1743282 of the Q735 produced du usually associated with art in which the ti thrown aside in a spirit of experimentatio experimented with new ways of seeing a nature of materials and functions of art. A Q1318295, which was characteristic for t Q673661 is characteristic of much model production is often called Q186030 or Q3

Modern art begins with the heritage of pa Q37693, Q34013 and Q82445 all of who development of modern art. At the begin Q5589 and several other young artists in Q153793, Q156272, Q45205, Q313096

the Darie art world with "wild" multi color

Nghệ thuật hiện đại (Q38166)

Bách khoa toàn thư mở Wikipedia

Thảo luận

Bài viết

Nghệ thuật hiện đại (Q1860: modern art) bao hàm các tác phẩm Q735 được sáng tác trong giai đoạn khoảng từ Q39844 đến Q35014, biểu thị các phong cách và Q5891 của nghệ thuật thời kỳ đó.^[1] Nghệ thuật hiện đại gạt sang một bên các truyền thống quá khứ với một tinh thần thể nghiệm.^[2] Các nghệ sĩ hiện đại thể nghiệm với những cách nhìn mới và với những ý tưởng mới về bản chất vật liệu và các chức năng của nghệ thuật. Nghệ thuật hiện đại có xu hướng xa rời Q1318295 (vốn dĩ là đặc trưng của nền nghệ thuật truyền thống) để hướng tới tính Q673661. Tác phẩm nghệ thuật ra đời trong các năm gần đây thường được gọi là Q186030 hoặc Q3318563.

Nghệ thuật hiện đại bắt đầu với di sản hội họa của nhiều họa sĩ như Q5582, Q35548, Q37693, Q34013 và Q82445, tất cả mạng



Modern art (Q38166)

Artistic works produced during the period ext

For other uses, see Q1205723. Not t

Modern art includes artistic work produc extending roughly from the 1860s to the and Q1743282 of the Q735 produced du usually associated with art in which the ti thrown aside in a spirit of experimentatio experimented with new ways of seeing a nature of materials and functions of art. A Q1318295. which was characteristic for Q673661 is characteristic of mucin mode production is often called Q186030 or Q3

Modern art begins with the heritage of page 1 Q37693, Q34013 and Q82445 all of who development of modern art. At the begin Q5589 and several other young artists in Q153793, Q156272, Q45205, Q313096 the Darie art world with "wild" multi color

Tìm kiếm Wikiped Sửa mã nguồn Bài viết Thảo luận Sửa đổi Thêm ~ Đọc

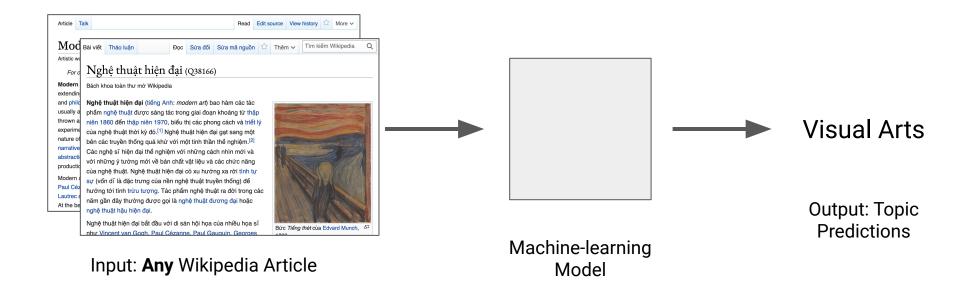
Nghệ thuật hiện đại (Q38166)

Bách khoa toàn thư mở Wikipedia

Nghệ thuật hiện đại (Q1860: modern art) bao hàm các tác phảm Q735 được sáng tác trong giai đoạn khoảng từ Q39844 đến Q35014, biểu thị các phong cách và Q5891 của nghệ thuật thời kỳ đó.^[1] Nghệ thuật hiện đại gạt sang một bên các truyền thống quá khứ với một tinh thần thể nghiệm. [2] Các nghệ sĩ hiện đại thể nghiệm với những cách nhìn mới và với những ý tưởng mơi về bản chất vật liệu và các chức năng của nghệ thuật. Nghệ thuật hiện đại có xu hướng xa rờ Q1318295 vốn dĩ là đặc trưng của nền nghệ thuật truyền thống) để hướng tới tính Q673661. Tác phẩm nghệ thuật ra đời trong các năm gần đây thường được gọi là Q186030 hoặc Q3318563.

Nghẹ thuật hiện đại bắt đầu với di sản hội họa của nhiều họa sĩ như Q5582, Q35546 Q37693, Q34013 và Q82445, tất cả mạng





Links + Newcomers: Add-a-link

A Multilingual Entity Linking System for Wikipedia with a Machine-in-the-Loop Approach

Martin Gerlach, Marshall Miller, Rita Ho, Kosta Harlan, Djellel Difallah. CIKM 2021. [arXiv]

Abstraction (disambiguation) Topics referred to by the same term. From Wikipedia, the free ency Abstraction is a process or result of generalization, removal Abstraction may also refer to: · Abstraction (art), art unconcerned with the literal depiction · Abstraction (computer science), a process of hiding detail Abstraction layer, an application of abstraction in comp · Hardware abstraction, an abstraction laver on top · Abstraction (linguistics), use of terms to concept · Abstraction (mathematics), a process of removing the · Hypostatic abstraction, a formal operation that trans · Lambda abstraction, a definition or an anonymous Abstraction (sociology), a pocess of considering sociole Nucleophilic abstraction, a micleophilic attack which Water abstraction, the process of taking water from any Abstracting electricity, the crime of diverting electricity are See also [edit source] Abstract (disambiguation) Abstracting service

· Abstract art, a movement in 20th-century Western art

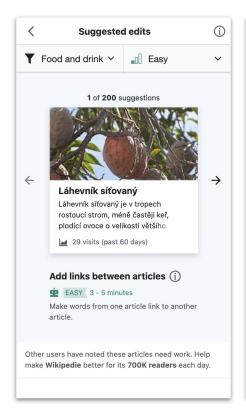
Modern art (Q38166)

Artistic works produced during the period extending roughly from the 1860s to the 1970s. Frq

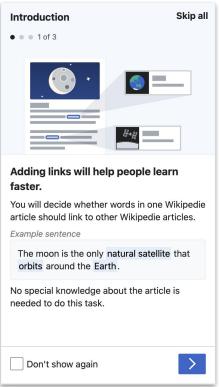
For other uses, see Modern art (disambiguation). Not to be confused with contem

Modern art includes artistic work produced during the period extending roughly from the 1860s to the 1970s, and denotes the styles and philosophies of the art produced during that era.^[1] The term is usually associated with art in which the traditions of the past have been thrown aside in a spirit of experimentation.^[2] Modern artists experimented with new ways of seeing and with fresh ideas about the nature of materials and functions of art. A tendency away from the narrative, which was characteristic for the traditional arts, toward abstraction is characteristic of much modern art. More recent artistic production is often called contemporary art or postmodern art.

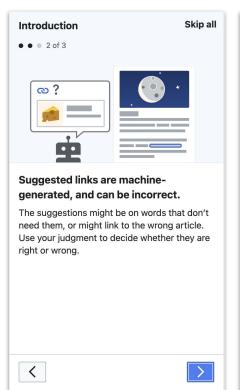
Choose a task



Onboarding: value



Onboarding: judgment

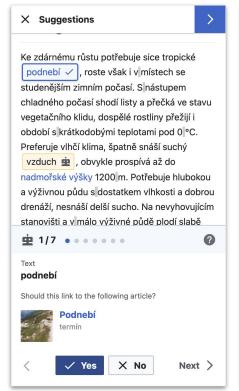


Onboarding: guidelines

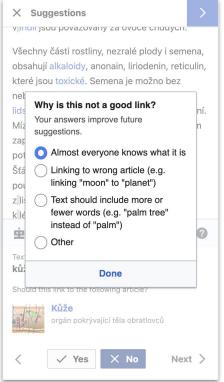


Learn more: https://www.mediawiki.org/wiki/Growth/Personalized_first_day/Structured_tasks/Add_a_link

Evaluate the suggestion



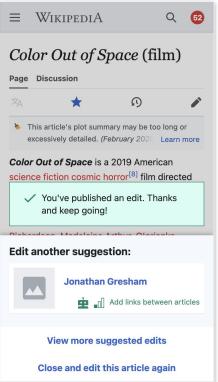
Feedback on algorithm



Edit summary



Next suggestion



Learn more: https://www.mediawiki.org/wiki/Growth/Personalized_first_day/Structured_tasks/Add_a_link

Takeaways

Overview: Links are awesome and powerful and (I think) underappreciated!

Success: Not all links are equal for readers -- some are much more likely to be clicked

Gender Gap: There are still large structural gaps in how well-incorporated articles about women are to the rest of Wikipedia.

Equitable Machine Learning: Even links that are never clicked can be valuable for building machine-learning models across many languages to aid in editing and other tasks.

Newcomers: Adding links is an excellent newcomer task and may one day help to address some of the structural biases in the coverage of links

Attribution

Team member photos (slide 2):

- o Pablo Aragón: https://commons.wikimedia.org/wiki/File:Pablo_Aragon.jpg (CC-BY-SA 4.0)
- Martin Gerlach: https://commons.wikimedia.org/wiki/File:Martin_Gerlach_20191209_151758.jpg (CC-BY-SA 4.0)
- o Isaac Johnson: https://commons.wikimedia.org/wiki/File:Isaac_Johnson.jpg (CC-BY-SA 4.0)
- Emily Lescak: https://commons.wikimedia.org/wiki/File:Emily_Lescak_landscape_portrait.jpg (CC-BY-SA 4.0)
- Miriam Redi: https://commons.wikimedia.org/wiki/File:Miriam_final-3.jpg (CC-BY-SA 4.0)
- Diego Sáez-Trumper: https://commons.wikimedia.org/wiki/File:Diego_S%C3%A1ez-Trumper.jpg (CC-BY-SA 4.0)
- Camera-shy: https://commons.wikimedia.org/wiki/File:Wikimedia_Foundation_office_camera_shy.jpg
 (CC-BY-SA 3.0)
- Leila Zia: https://commons.wikimedia.org/wiki/File:Zia,_Leila_Feb_2014.jpg (CC-BY-SA 3.0)

Wikipedia Screenshots

- https://en.wikipedia.org/wiki/Modern_art (CC-BY-SA 3.0) (first on slide 3)
 - Image in en:Modern Art of <u>Van Gogh Country road in Provence by night</u> (<u>Public Domain</u>)
- https://vi.wikipedia.org/wiki/Ngh%E1%BB%87_thu%E1%BA%ADt_hi%E1%BB%87n_%C4%91%E1%BA%A1i (CC-BY-SA 3.0) (first on slide 19)
 - Image in vi:Nghệ_thuật_hiện_đại of <u>Edvard Munch</u>, 1893, <u>The Scream</u>, oil, tempera and <u>pastel on cardboard</u>, 91 x 73 cm, <u>National Gallery of Norway</u> (<u>Public Domain</u>)

- Link success page diagram (slide 6): https://arxiv.org/pdf/1611.02508.pdf#page=2
 (CC-BY 4.0)
- WikiNav Clickstream Screenshots (slides 8, 9)
 - Freely licensed per: https://github.com/mnzpk/wikinav/blob/main/README.md#license
- Figure 5. PageRank (slide 12):
 https://epjdatascience.springeropen.com/articles/10.1140/epjds/s13688-016-0066-4/figures/5 (CC-BY 4.0)
- Incoming Links Graph (slide 13):
 https://figshare.com/articles/presentation/Bridging_Wikipedia_s_Gender_Gap_Quantify_ing_and_Assessing_the_Impact_of_Two_Feminist_Interventions_-_Wikimedia_Research_Showcase_-_July_2021/15059532 (CC-BY 4.0)
- Add-a-link prototypes (slides 28, 29) copied from
 <u>https://commons.wikimedia.org/wiki/File:Wikimania21-Editing-with-machine-learning-Link-recommendation.pdf</u> (CC-BY-SA 4.0)
 - Wikipedia articles shown within: https://cs.wikipedia.org/wiki/L%C3%A1hevn%C3%ADk_s%C3%AD%C5%A5ovan%C3%BD (CC-BY-SA 3.0) and https://en.wikipedia.org/wiki/Color_Out_of_Space_(film) (CC-BY-SA 3.0)