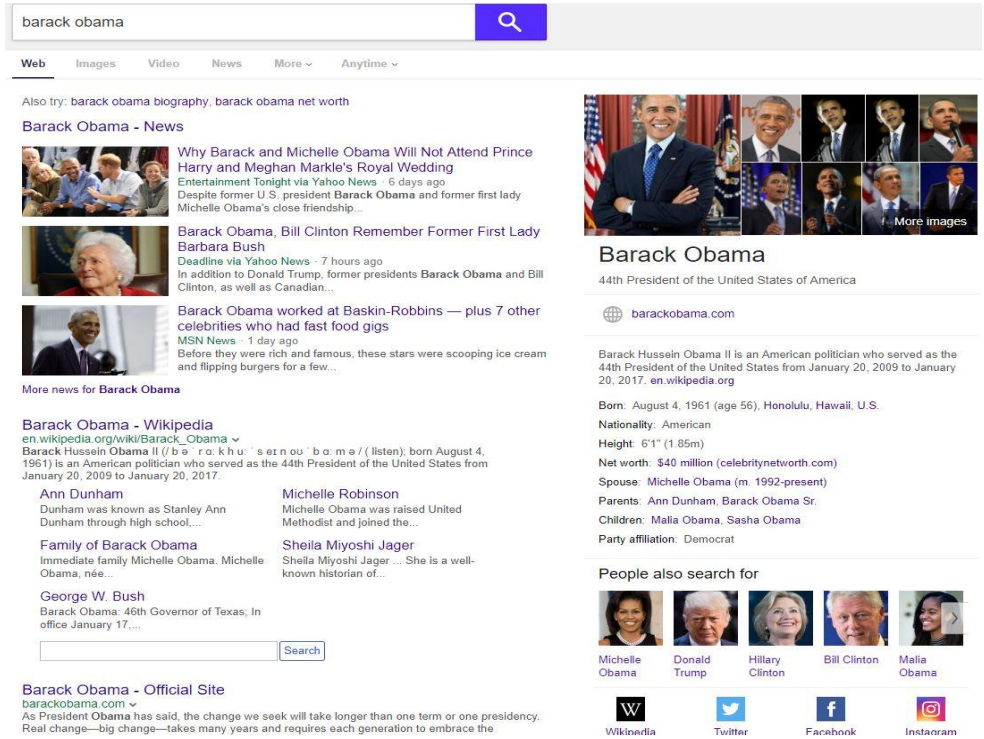
Automatic Trivia Fact Extraction from Wikipedia

Qiancheng Li, Aniket Bonde
- Under Supervision of Prof. Ruihong Huang

Motivation

- Web Search is now an exploratory activity
- Improving engagement is a key goal
- Most queries are entity related



The Contribution

Barack Obama

44th U.S. President

barackobama.com

Barack Hussein Obama II is an American politician who served as the 44th President of the United States from January 20, 2009 to January 20, 2017. Wikipedia

Born: August 4, 1961 (age 56 years), Kapiolani Medical Center for Women and Children, Honolulu, HI

Height: 6′ 1″

Presidential term: January 20, 2009 – January 20, 2017

Education: Harvard Law School (1988–1991), MORE

Parents: Ann Dunham, Barack Obama Sr.

Did you know? Grammy Award winner

Wikipedia Category

- Natural Language - obstacle of detecting trivia
- Wikipedia also has structured information
- **Categories:** set of articles with a shared topic

Categories: Barack Obama | 1961 births | 20th-century American writers | 20th-century scholars | 21st-century American politicians | 21st-century American writers | 21st-century scholars | Activists from Illinois | African-American feminists | African-American non-fiction writers | American Christians | American Protestants | American non-fiction writers | African-American people in Illinois politics | African-American United States presidential candidates | African-American United States Senators | American civil rights lawyers | American community activists | American feminist writers | American feminists | American legal scholars | American Nobel laureates | American people of English descent | American people of French descent | American people of German descent | American people of Irish descent | American people of Luo descent | American people of Scottish descent | American people of Swiss descent | American people of Welsh descent | American political writers | American politicians of Luo descent | Columbia University alumni | Democratic Party (United States) presidential nominees | Democratic Party Presidents of the United States | Democratic Party United States Senators | Grammy Award winners | Harvard Law School alumni | Illinois Democrats | Illinois lawyers | Illinois State Senators | Living people | Male feminists | Nobel Peace Prize laureates | Obama family | Occidental College alumni | Politicians from Chicago | Politicians from Honolulu | Presidents of the United States | Punahou School alumni | United States presidential candidates, 2008 | United States presidential candidates, 2012 | United States Senators from Illinois | University of Chicago Law School faculty | Writers from Chicago

Wikipedia Category

Categories: Barack Obama | 1961 births | 20th-century American writers | 20th-century scholars | 21st-century American politicians | 21st-century American writers | 21st-century scholars | Activists from Illinois | African-American feminists | African-American non-fiction writers | American Christians | American Protestants | American non-fiction writers | African-American people in Illinois politics | African-American United States presidential candidates | African-American United States Senators | American civil rights lawyers | American community activists | American feminist writers | American feminists | American legal scholars | American Nobel laureates | American people of English descent | American people of French descent | American people of German descent | American people of Irish descent | American people of Luo descent | American people of Scottish descent | American people of Swiss descent | American people of Welsh descent | American political writers | American politicians of Luo descent | Columbia University alumni | Democratic Party (United States) presidential nominees | Democratic Party Presidents of the United States | Democratic Party United States Senators | Grammy Award winners | Harvard Law School alumni | Illinois Democrats | Illinois lawyers | Illinois State Senators | Living people | Male feminists | Nobel Peace Prize laureates | Obama family | Occidental College alumni | Politicians from Chicago | Politicians from Honolulu | Presidents of the United States | Punahou School alumni | United States presidential candidates, 2008 | United States presidential candidates, 2012 | United States Senators from Illinois | University of Chicago Law School faculty | Writers from Chicago

- Rank these categories by how trivia-worthy they are
- **Challenge:** Formalize notion of trivia-worthy

How a 23-year-old makes $500,000 a year tweeting random facts

Madeline Stone

Feb. 12, 2015, 2:10 PM 183,522

FACEBOOK LINKEDIN TWITTER EMAIL COPY LINK

When Kris Sanchez joined Twitter in 2009, he didn't expect much to come of it.

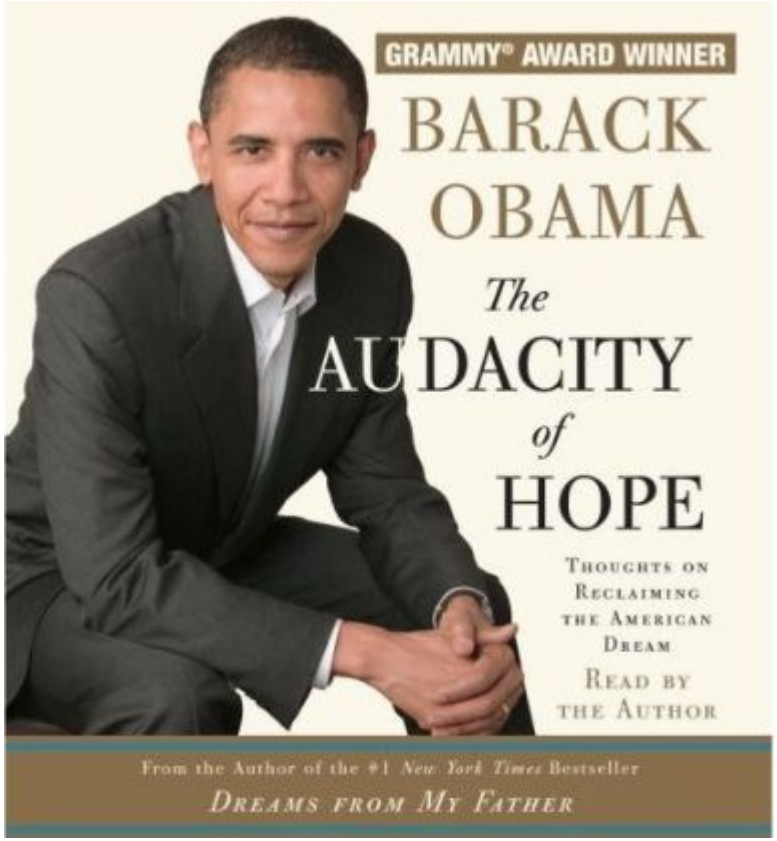
"I really started my Twitter account because I wanted to follow Britney Spears," Sanchez told Business Insider. "I'm a huge fan."

He found he didn't have much to tweet about in his daily life, so he started sharing random facts he found while procrastinating on the internet.

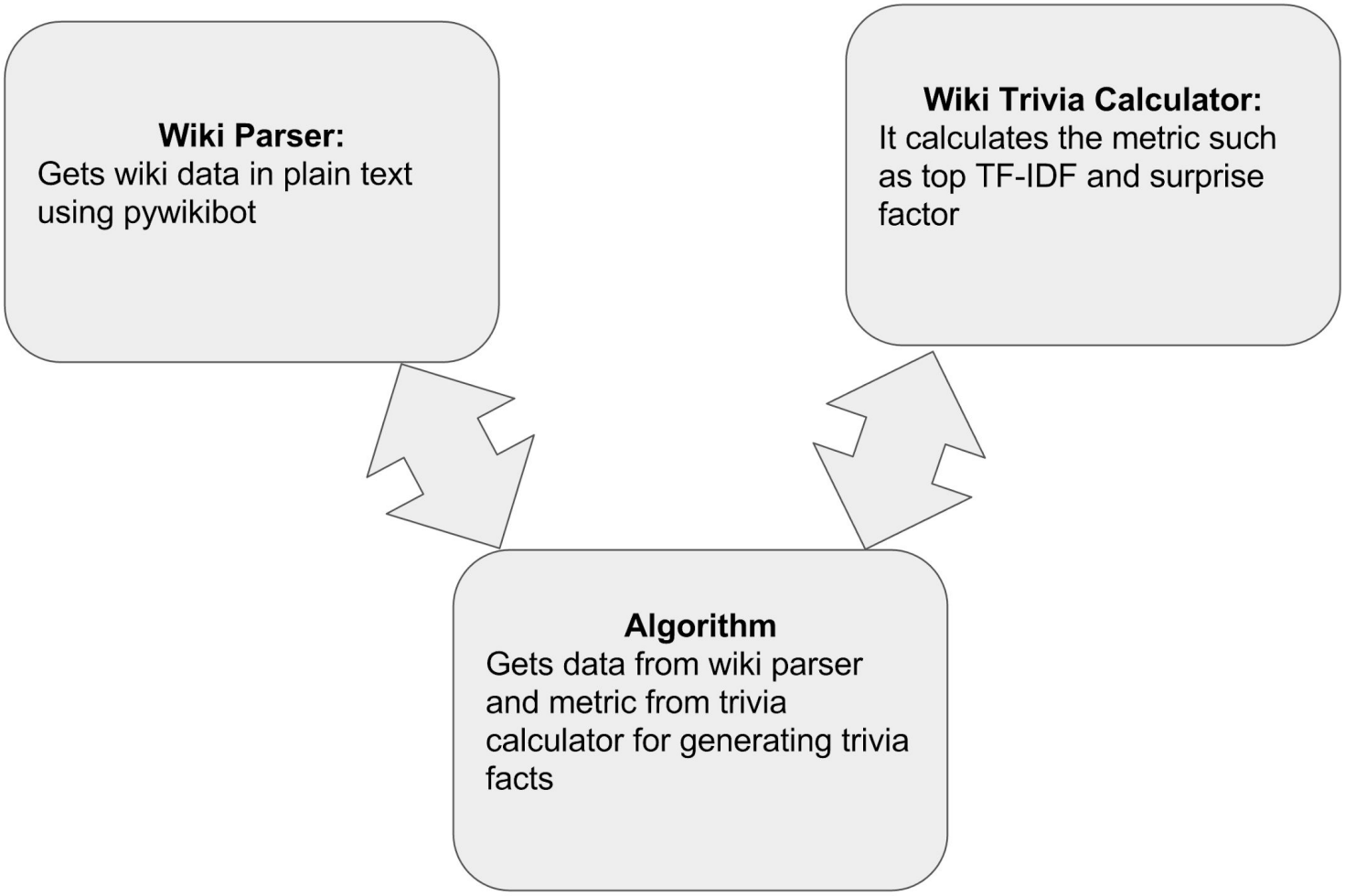
Kris Sanchez and his UberFacts team. Eric Charbonneau

Not All Facts Are Trivia

- Obama was a US president
- Obama was born in 1961
- Obama won a Grammy award



Architecture



Trivia Worthy

- Surprise: get people's attention

$$\sigma(a, C) = \frac{1}{|C| - 1} \sum_{a \neq a' \in C} \sigma(a, a') \qquad surp(a, C) = \frac{1}{\sigma(a, C)}$$

- Cohesiveness

$$cohesive(C) = \frac{1}{\binom{|C|}{2}} \sum_{a \neq a'} \sigma(a, a')$$

- Trivia Worthy

$$trivia(a, C) = cohesive(C) \cdot surp(a, C)$$

Top Surprise:

20th-century Austrian people

Women in technology

Radio pioneers

American anti-fascists

American people of Hungarian-Jewish descent

Top Cohesiveness:

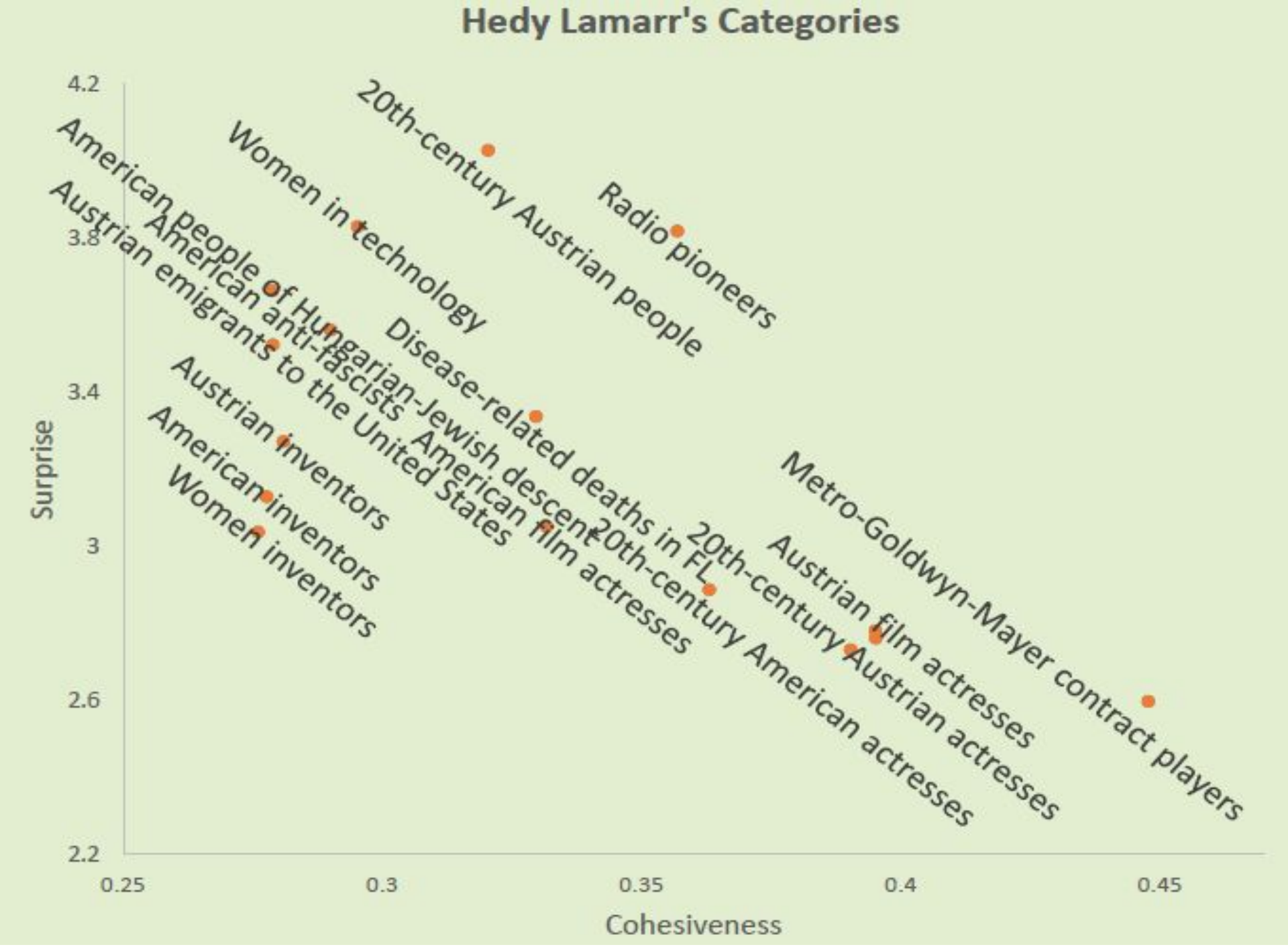
Metro-Goldwyn-Mayer contract players

Actresses from Vienna

Austrian film actresses

20th-century Austrian actresses

American film actresses


Hedy Lamarr's Categories
Surprise
Cohesiveness
4.2
3.8
3.4
3
2.6
2.2
0.25
0.3
0.35
0.4
0.45
20th-century Austrian people
Radio pioneers
Women in technology
American people of Hungarian-Jewish descent
American anti-fascists
Austrian emigrants to the United States
Disease-related deaths in FL
American film actresses
20th-century Austrian actresses
Austrian film actresses
Metro-Goldwyn-Mayer contract players
Austrian inventors
American inventors
Women inventors
20th-century American actresses

Algorithm

Algorithm 4 Article Similarity

1: **function** ARTICLESIM($article1, article2$)
2: $K = 10$
3: $T1 = TopTFIDF(article1, K)$
4: $T2 = TopTFIDF(article2, K)$
5: $sim = (article1, article2)$
return sim

Algorithm

Algorithm 3 Cohesiveness

```
1: function COHESIVENESS(category)
2:     sum, count = 0
3:     for every article pair a1 ≠ a2 in category C do
4:         sim = ArticleSim(a1, a2)
5:         sum = sum + sim
6:         count = count + 1
7:     cohesiveness = sum/count
   return cohesiveness
```

Algorithm

Algorithm 2 Surprise

```
1: function SURPRISE(inputArticle, C)
2:     sum, count = 0
3:     for every article a ≠ inputArticle in category C do
4:         sim = ArticleSim(inputArticle, a)
5:         sum = sum + sim
6:         count = count + 1
7:     similarityToCategory = sum/count
8:     surprise = 1/similarityToCategory
   return surprise
```

Algorithm

Algorithm 1 Trivia Extract

```
1: function TRIVIAEXTRACT(inputArticle)
2:     for every category C of inputArticle do
3:         surp = Surprise(inputArticle, C)
4:         cohes = Cohesiveness(C)
5:         C.trivia = cohes . surp
   return category C with maximum trivia score
```

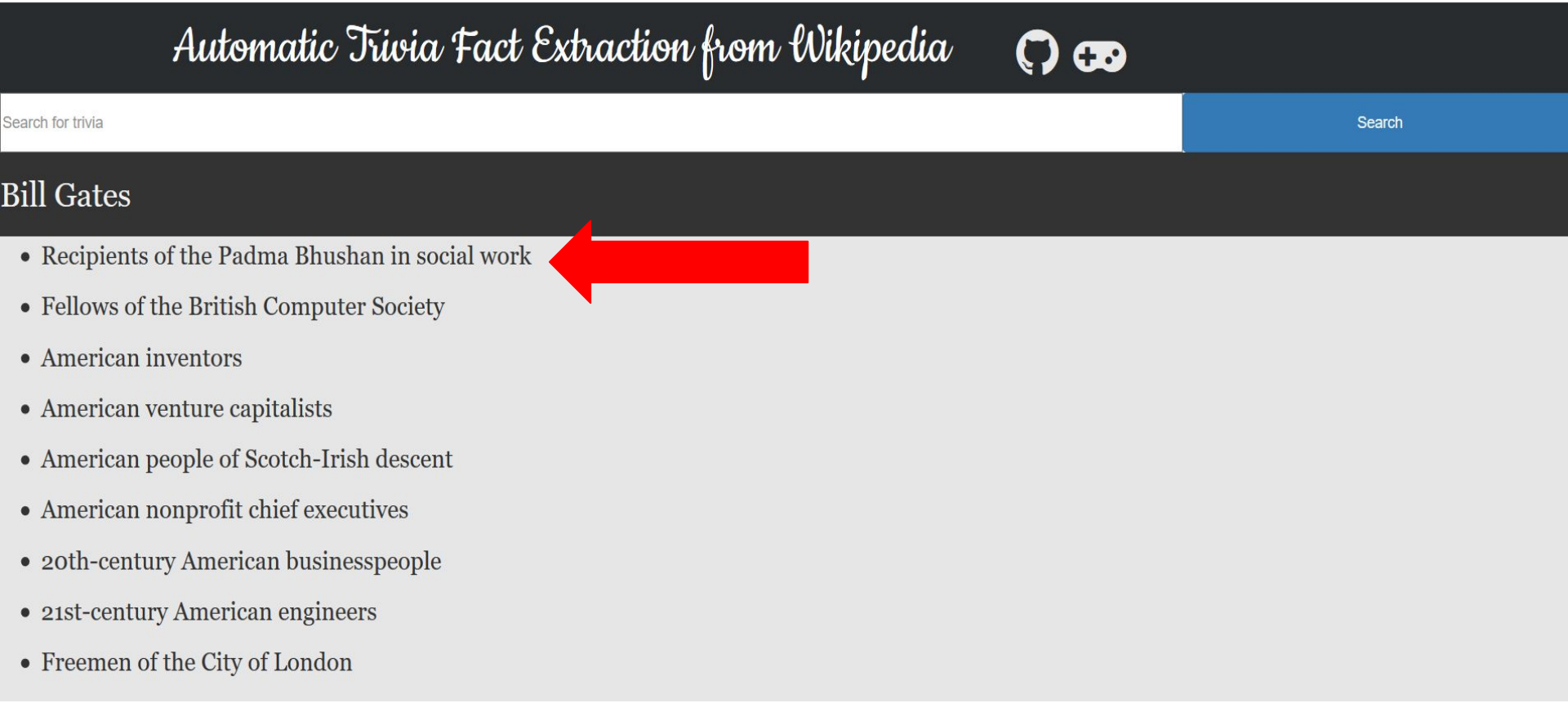
Result: Barack Obama for example

Punahou School alumni	1.3387910646917047
Grammy Award winners	1.2859668279303453
American Nobel laureates	1.2565331843053882
Nobel Peace Prize laureates	1.1786827577086749
American feminist writers	1.1551973277907599
African-American feminists	1.1167987835685782
American feminists	1.0896562059932569
21st-century American politicians	1.0057797842693141
Democratic Party United States Senators	0.99517280067274383
Harvard Law School alumni	0.994574666360582

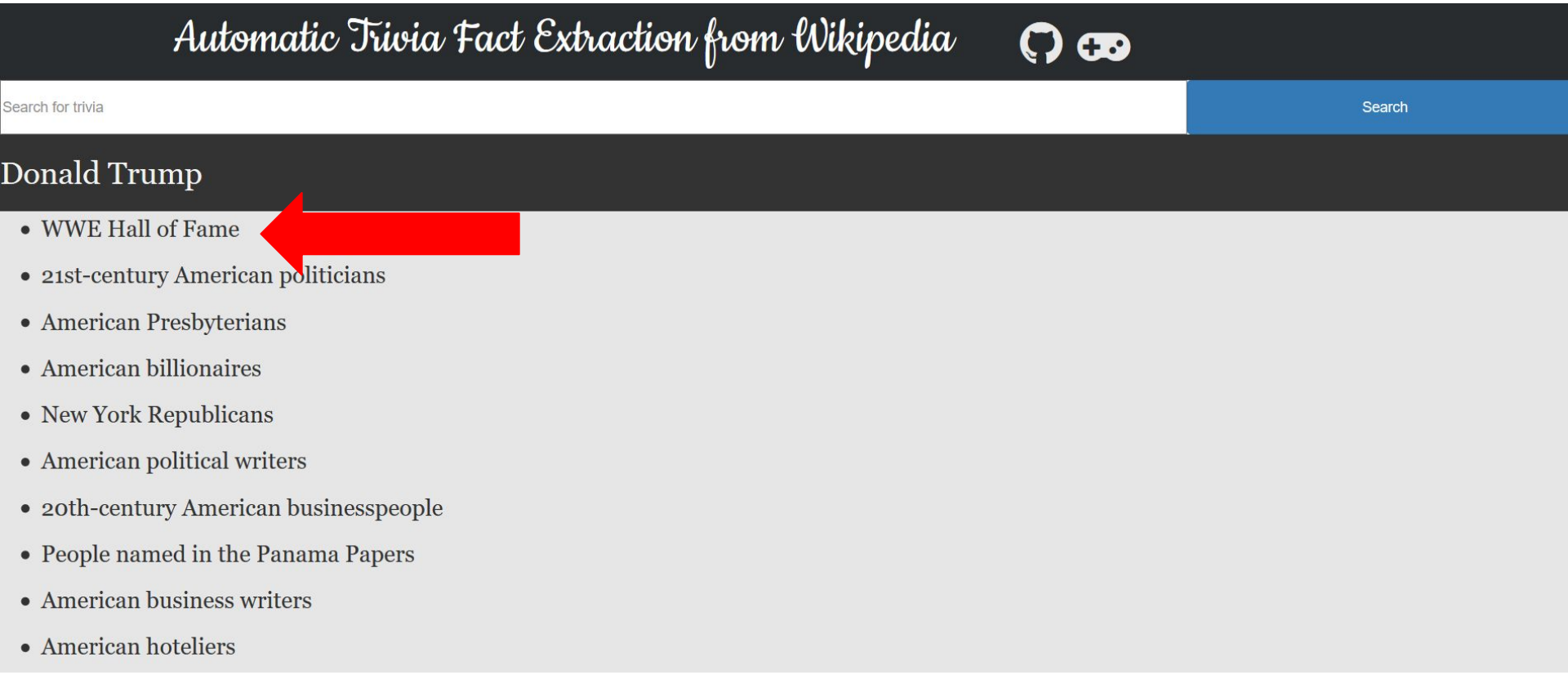
Evaluation

- Wikipedia Trivia Miner (WTM, IJCAI ‘15)
- Compared top 5 (Didn’t get 100% overlap)
- Due to randomly sampling ‘k’ entities for a category
- Best Way - Run user studies (Not done due to lack of time/resources)

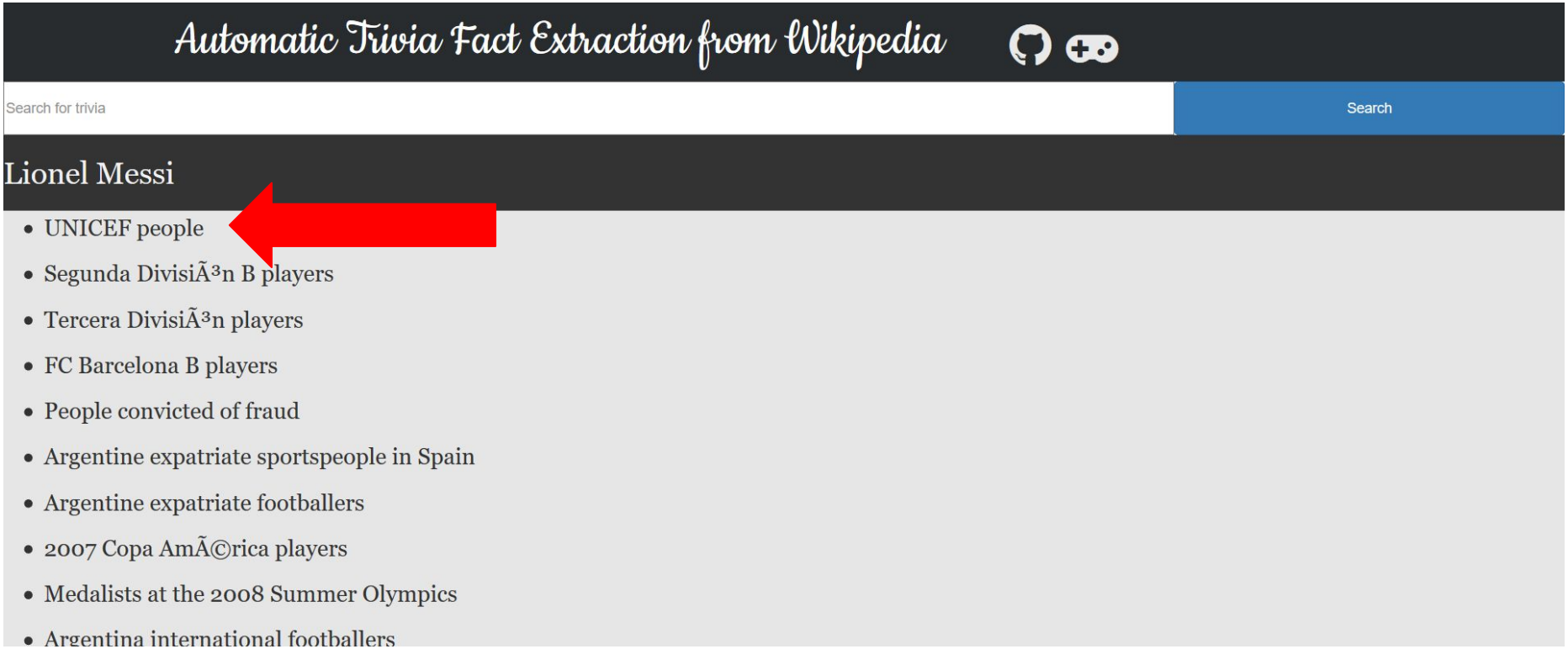
App!


Automatic Trivia Fact Extraction from Wikipedia
Search for trivia
Search
Bill Gates
Recipients of the Padma Bhushan in social work
Fellows of the British Computer Society
American inventors
American venture capitalists
American people of Scotch-Irish descent
American nonprofit chief executives
20th-century American businesspeople
21st-century American engineers
Freemen of the City of London

App!


Automatic Trivia Fact Extraction from Wikipedia
Search for trivia
Search
Donald Trump
WWE Hall of Fame
21st-century American politicians
American Presbyterians
American billionaires
New York Republicans
American political writers
20th-century American businesspeople
People named in the Panama Papers
American business writers
American hoteliers

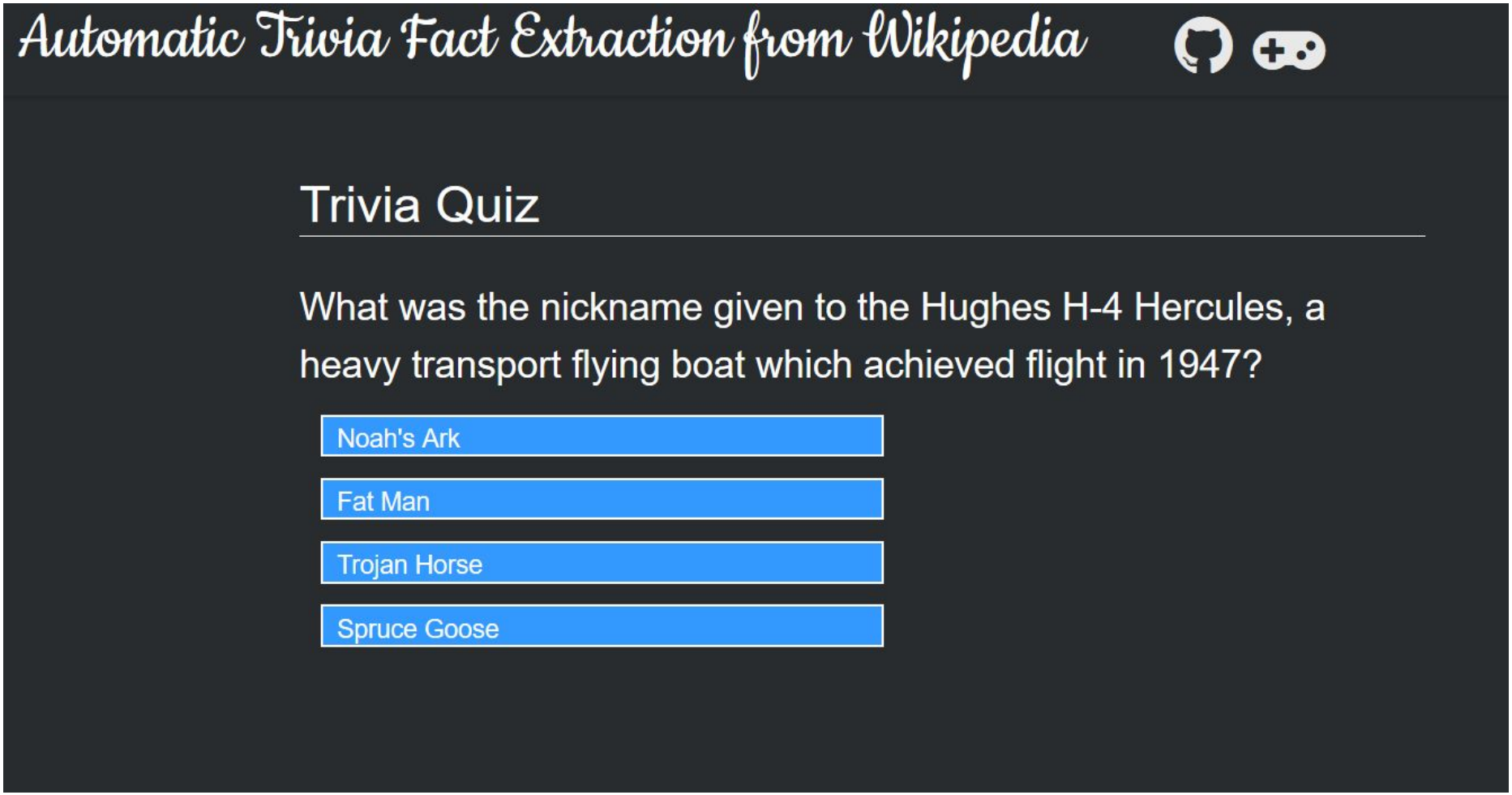
App!


Automatic Trivia Fact Extraction from Wikipedia
Search for trivia
Search
Lionel Messi
UNICEF people
Segunda DivisiÃ³n B players
Tercera DivisiÃ³n players
FC Barcelona B players
People convicted of fraud
Argentine expatriate sportspeople in Spain
Argentine expatriate footballers
2007 Copa AmÃ©rica players
Medalists at the 2008 Summer Olympics
Argentina international footballers

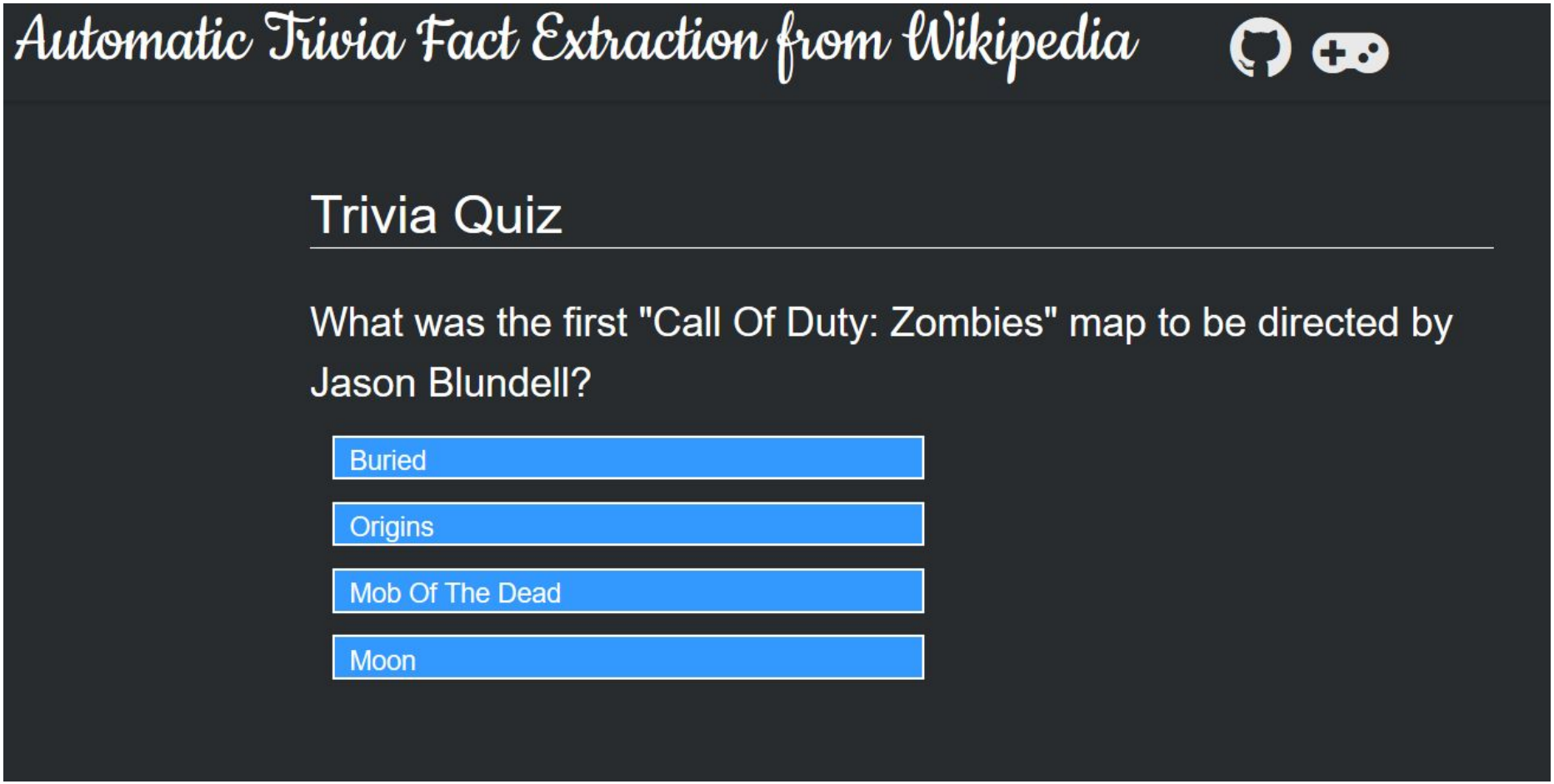
App!

Automatic Trivia Fact Extraction from Wikipedia
Trivia Quiz
Select Game Type
Quiz
Learning Game

App!


Automatic Trivia Fact Extraction from Wikipedia
Trivia Quiz
What was the nickname given to the Hughes H-4 Hercules, a heavy transport flying boat which achieved flight in 1947?
Noah's Ark
Fat Man
Trojan Horse
Spruce Goose

App!


Automatic Trivia Fact Extraction from Wikipedia
Trivia Quiz
What was the first "Call Of Duty: Zombies" map to be directed by Jason Blundell?
Buried
Origins
Mob Of The Dead
Moon

Conclusion

- Detect good trivia
- Introduced formulation: Surprise, cohesiveness
- Increase user engagement

References

Mika, Peter. "Entity search on the web." Proceedings of the 22nd International Conference on World Wide Web. ACM, 2013.

Tsurel, D., Pelleg, D., Guy, I., Shahaf, D.: Fun facts: automatic trivia fact extraction from wikipedia (2016). arXiv preprint arXiv:1612.03896

Yin, Xiaoxin, and Sarthak Shah. "Building taxonomy of web search intents for name entity queries." Proceedings of the 19th international conference on World wide web. ACM, 2010.

Sergey Chernov, Tereza Iofciu, Wolfgang Nejdl, and Xuan Zhou. Extracting semantics relationships between Wikipedia categories. SemWiki, 206, 2006.

Abhay Prakash, Manoj K. Chinnakotla, Dhaval Patel, and Puneet Garg. Did you know?: Mining interesting trivia for entities from wikipedia. In Proceedings of the 24th International Conference on Artificial Intelligence, IJCAI'15, pages 3164--3170. AAAI Press, 2015.

References

Matthew Merzbacher. Automatic generation of trivia questions. In International Symposium on Methodologies for Intelligent Systems, pages 123-130.Springer, 2002.

Marco Baroni, Georgiana Dinu, and Germ´an Kruszewski. Don't count, predict! a systematic comparison of context-counting vs. context-predicting semantic vectors. In Proceedings of Association for Computational Linguistics (ACL), volume 1, 2014.

Tom Kenter and Maarten de Rijke. Short text similarity with word embeddings. In Proceedings of the 4th ACM International on Conference on 353 Information and Knowledge Management, CIKM '15, pages 1411–1420, New York, NY, USA, 2015. ACM.

Tom Kenter and Maarten de Rijke. Short text similarity with word embeddings. In Proceedings of the 4th ACM International on Conference on 353 Information and Knowledge Management, CIKM '15, pages 1411–1420, New York, NY, USA, 2015. ACM.