

Analysis of Evaluation Results of SPARQL Queries (All Problematic Cases)

FALSE POSITIVES:

Main reasons for occurrence:

1. Disambiguation Error (in mapping Persons)

- i. For the query “**#3 African American Film Producer**”, an article contains an incorrect mapping of Michael Jackson to some Michael J. Jackson, a Mercedes-Benz executive.
- ii. For the query “Ferrari Formula One drivers”, an article is present which mentions Tony Brooks, who is an NFL American footballer but has been incorrectly mapped to Formula One driver of the same name.
- iii. For the query “**#9 Australian cricketers One Day Internationals**”, results contain articles in which there is incorrect mapping of Kevin Wright, Alan Connolly and Matthew Elliott, all Australian cricketers but in the articles referring to some other person of the same name. In another article, the name “John Dyson” was incorrectly linked to the Australian cricketer John Dyson, but referred to John Dyson, the deputy mayor of New York at the time of Rudolph Giuliani’s mayoralty.
- iv. For the query “**#19 NASA civilian astronauts**”, an article contains an incorrect mapping of Gregory Johnson, a NASA civilian astronaut to Greg Johnson, president of Franklin Templeton investments.

2. Disambiguation Error (caused by Abbreviations)

- i. For the query “**#11 Video Game Consoles**”, two of the results contain a wrong mapping of the company New England Electric System (NES) to the Nintendo Game console NES. Similarly, another result contains incorrect mapping of company Weatherford International (WII) to Nintendo Game Console Wii.

3. Disambiguation Error (for sentences with comma as separator)

- i. For the query “**#16 Indian meat dishes**”, two of the articles contain sentences in which chicken and butter are separated by a comma like “.....butter, chicken....” but they get considered by the entity annotation system combined together and get incorrectly mapped to the famous Indian meat dish of Butter Chicken.

4. Exceptional Cases

- i. For the query “**#12 Indian Padma Shri recipients**”, four of the results contain articles mentioning Ben Kingsley, the famous English actor who has won Academy, Golden Globe and Oscar Awards. Padma Shri is awarded to citizens of India in recognition of their distinguished contribution in various spheres of activity including the Arts, Education, Industry, Literature, Science, Sports, Medicine, Social

Service and Public Affairs. But in some exceptional cases, it gets also awarded to distinguished individuals who were not citizens of India but did contribute in various ways to India. Ben Kingsley, is a British Citizen but has parents who immigrated from India to the UK. Him getting awarded Padma Shri was a special case. But since he is not an Indian citizen, we classified his case as a False Positive.

FALSE NEGATIVES:

Main reasons for occurrence:

1. Entity Extraction System not being able to disambiguate correctly

- i. For the query **"#2 Los Angeles Lakers player"**, both Google and HistDiv contain a result mentioning Mike Johnson, a Los Angeles Lakers player. Mike Johnson was commonly known and is mentioned on DBpedia as Magic Johnson, and hence him getting mentioned as Mike Johnson leads to him not being able to get disambiguated by the entity extraction system.

2. Entity Extraction System did not manage to recognize one of the entities of interest

- i. For the query **"#16 Indian meat dishes"**, a HistDiv result mentions "Tandoori Murgh" which is the same as Tandoori Chicken as "Murgh" is the Hindi reference for "Chicken", and there is a DBpedia page mentioning the same dish by Tandoori Chicken. So, the entity extraction system does not recognize "Tandoori Murgh" as a reference to "Tandoori Chicken".
- ii. For the query **"#8 Boeing 747 aircraft accidents"**, three of the results by HistDiv contain a vague reference of an accident involving the Boeing 747 but do not clearly point to what the incident was. Some of the lines are *"The F.A.A. formed the group after the Iowa crash and said it was also in response to incidents several years ago involving a Boeing 747 and a Lockheed L-1011 in which engine failures caused damage to aircraft control mechanisms."* and *"The other accidents involved a Japan Air Lines Boeing 747, which crashed in 1985, killing 516 people, after an internal wall ruptured, and an Eastern Airlines Lockheed L-1011, which landed safely at Kennedy Airport in 1981 after its tail engine broke apart."* These two lines indicate some mishappening with Boeing 747 aircrafts, but do not clearly indicate which incident like Pan Am Flight 103 or Air India Flight 102 and hence do not get detected by the Entity Extraction System.