

Role(Models): Extract Character Roles And Genders from Movie Summaries

Wenqin Chen, Jordan Shedlock

Project Overview

Problem Statement:

The representation of character roles and genders in movies has significant social impact. To study the distribution of roles and genders in movies, we need to first get the data. How can we effectively extract characters' roles and genders from movie summaries?

Dataset:

Role and summary information for around 34,000 US movies with 10 or more IMDb reviews and IMDb and/or Wikipedia summaries.

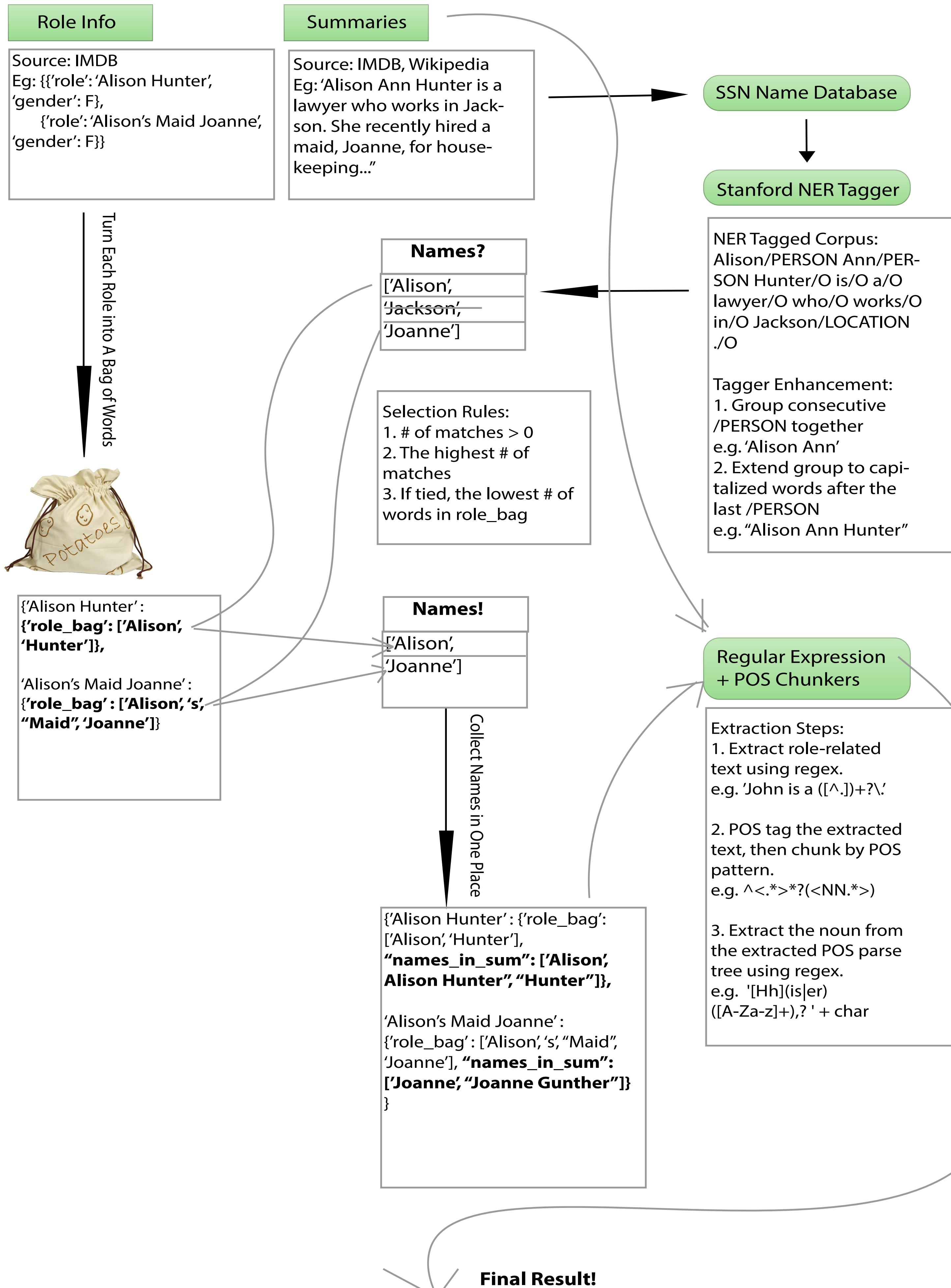
Data Sources:

- IMDb: Plain text role, summary, and country data from IMDb's Alternative Interfaces
- Wikipedia: Summaries of around 42,000 movies collected by David Bamman, Brendan O'Connor, and Noah Smith
- Social Security Administration: List of first names since the 1880s
- CMU Artificial Intelligence Repository: List of surnames collected by Mark Kantrowitz

Project Code:

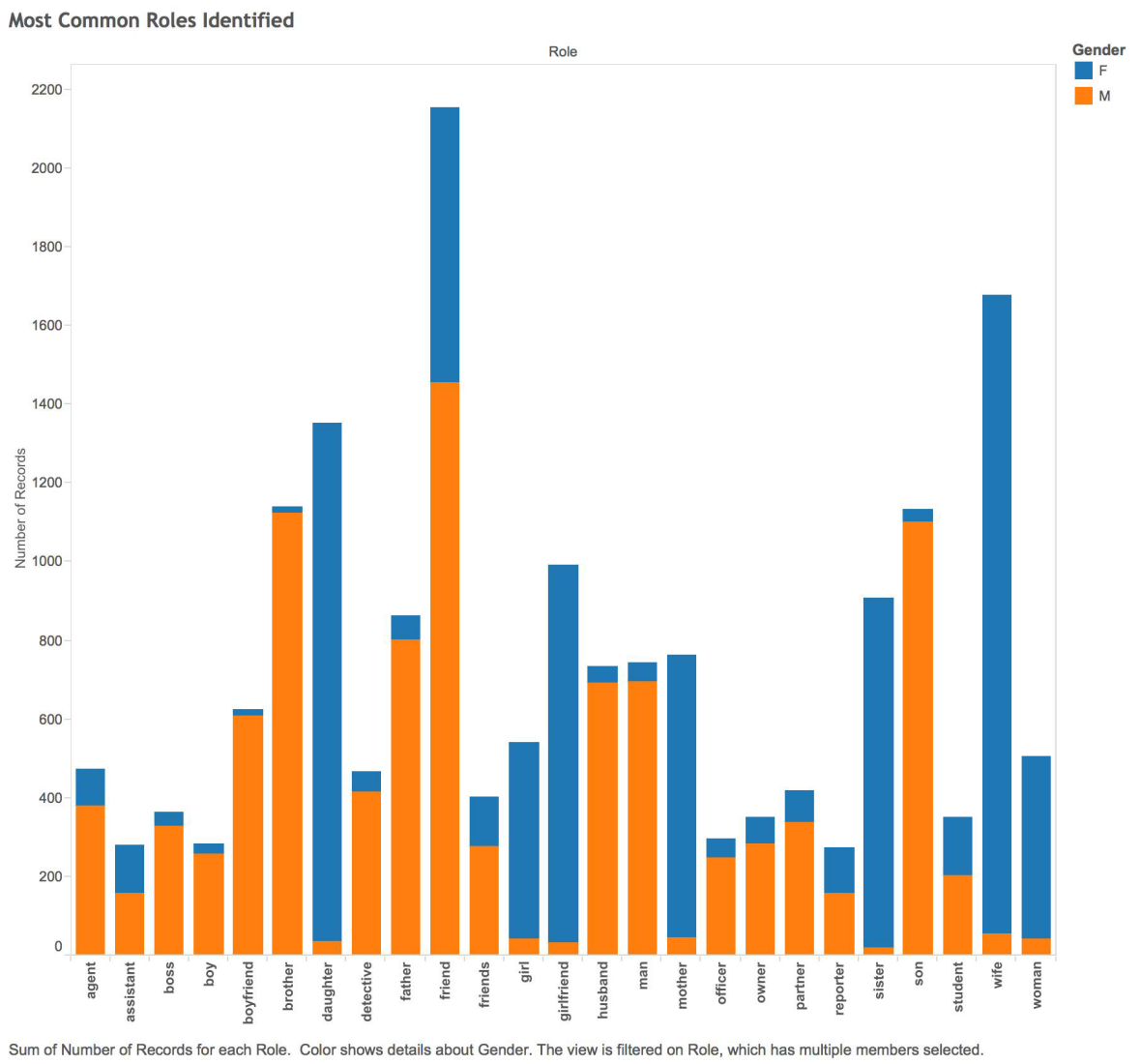
https://github.com/jwqchen/NLP_movie_roles

Project Flowchart:

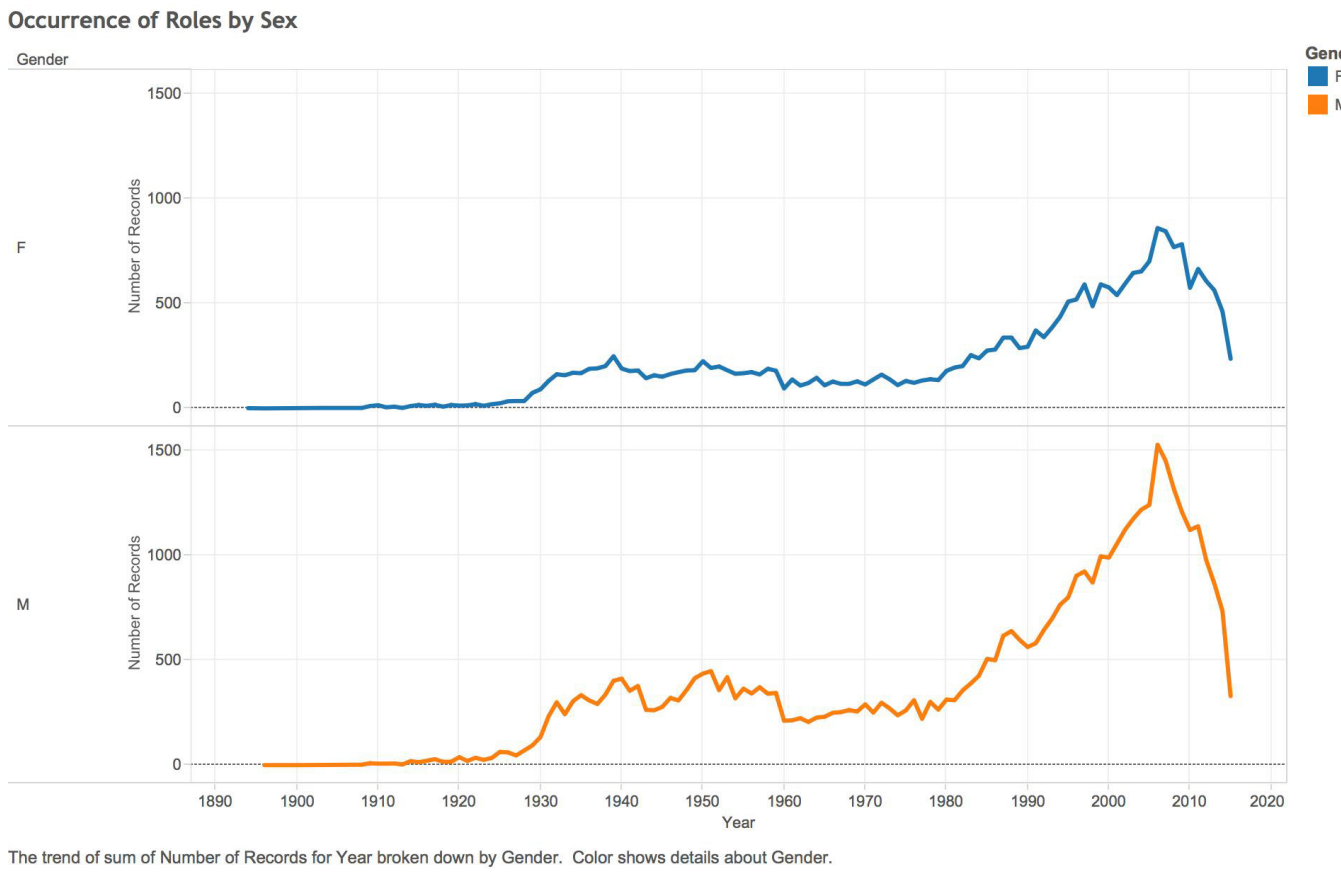


Findings & Analysis

We identified **71,216** descriptive phrases for **114,922** roles. The most common ones are shown below, divided by sex. In the case of sex-specific roles (e.g. family relationships), most of these phrases were associated with the correct sex. Although most of the top roles are related to family or relationships, most of the few professional roles that do appear (officer, owner, agent, detective) skew male.



Male and female roles seem to display the same trends (which may be related to the completeness of the data); however, male roles consistently account for the higher share.



Most Common Roles

The most common roles for both sexes are family roles.

| Men | Women |
|---|----------------|
| 1. Friend | 1. Wife |
| 2. Brother | 2. Daughter |
| 3. Son | 3. Girlfriend |
| 4. Father | 4. Sister |
| 5. Man | 5. Mother |
| 6. Husband | 6. Friend |
| 7. Boyfriend | 7. Girl |
| But further down the list, the roles diverge sharply. | |
| 8. Detective | 8. Woman |
| 9. Agent | 9. Actress |
| 10. Partner | 10. Singer |
| 11. Boss | 11. Student |
| 12. Owner | 12. Secretary |
| 13. Friends | 13. Widow |
| 14. Boy | 14. Niece |
| 15. Officer | 15. Friends |
| 16. Captain | 16. Assistant |
| 17. Leader | 17. Nurse |
| 18. Director | 18. Teacher |
| 19. Lawyer | 19. Reporter |
| 20. Student | 20. Sweetheart |
| 21. Professor | 21. John |
| 22. Manager | 22. Aunt |
| 23. Attorney | 23. Agent |
| 24. Uncle | 24. Neighbor |
| 25. Reporter | 25. Star |
| 26. Assistant | 26. Cousin |
| 27. Cousin | 27. Teenager |
| 28. Lieutenant | 28. Waitress |
| 29. Producer | 29. Partner |
| 30. Businessman | 30. Mistress |