**Homework 2 Report**

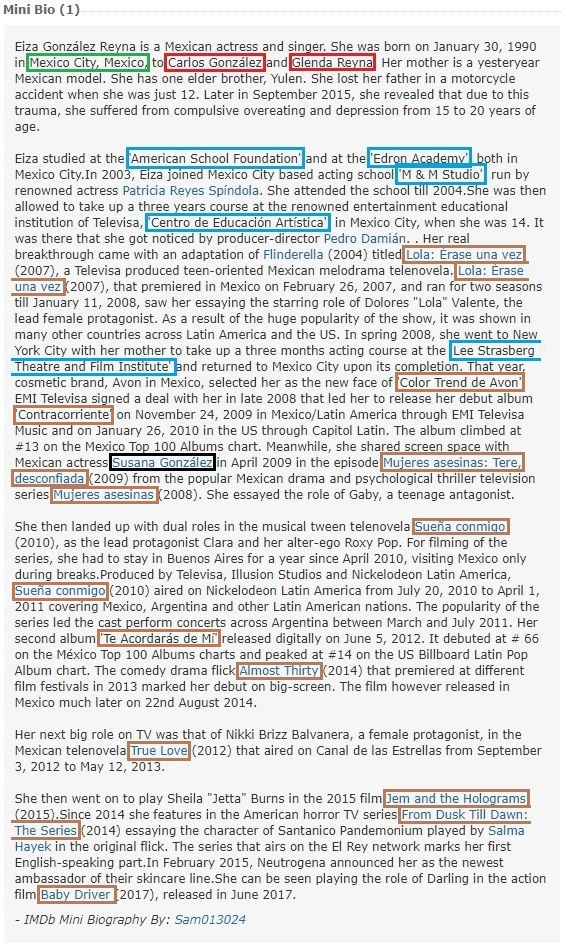
**Matheus Schmitz**

**Task 1:**

Date used: January 30th

URL: https://www.imdb.com/search/name/?birth\_monthday=01-30

**Task 2.1:**





**Task 2.2:**

birthplace\_lexical = [

    {'LOWER': 'born'},

    {'OP': '\*'},

    {'LOWER': 'in'},

    {'TEXT': {'REGEX': '\s\*'}, 'OP': '\*'},

    {'IS\_PUNCT': True, 'OP': '\*'},

    {'TEXT': {'REGEX': '\s\*'}, 'OP': '+'},

    {'IS\_PUNCT': True, 'OP': '\*'},

    ]

education\_lexical = [

    {'TEXT': {'REGEX': '^(attend|attended|went|studied|University|College|trained|graduated)$'}},

    {'OP': '+'},

    ]

parents\_lexical = [

    {'LOWER': {'REGEX': '^(born|son|daughter|parents)$'}},

    {'OP': '\*'},

    {'LOWER': {'REGEX': '^(to|of|are)$'}},

    {'TEXT': {'REGEX': '\s\*'}, 'OP': '\*'},

    {'IS\_PUNCT': True, 'OP': '\*'},

    {'TEXT': {'REGEX': '\s\*'}, 'OP': '+'},

    {'IS\_PUNCT': True, 'OP': '\*'},

    {'LOWER': 'and','OP': '?'},

    {'TEXT': {'REGEX': '\s\*'}, 'OP': '\*'},

    {'IS\_PUNCT': True, 'OP': '\*'},

    {'TEXT': {'REGEX': '\s\*'}, 'OP': '+'},

    {'IS\_PUNCT': True, 'OP': '\*'},

    ]

awards\_lexical = [

    {'LOWER': {'REGEX': '^(winner|recipient|won|awarded|award-winning|nominated|nominee|nomination)$'}},

    {'LOWER': {'REGEX': '^(of|the|on|for|as)$'}},

    {'OP': '+'}

    ]

performances\_lexical = [

    {'LOWER': {'REGEX': '^(star|starred|known|appeared|appearing|appearances|debute|album|recorded|play|played|role)$'}},

    {'LOWER': {'REGEX': '^(in|for|on|for|as)$'}},

    {'OP': '+'},

    ]

colleagues\_lexical = [

    {'IS\_TITLE': '+', 'OP': '\*'},

    {'LOWER': {'REGEX': '^(starring|collaborating|working|worked|partnering|sharing.\*screen.\*space|appears.\*opposite|opposite)$'}},

    {'LOWER': {'REGEX': '^(with|alongside|to)$'}}, {'OP': '\*'},

    {'OP': '\*'},

    ]

birthplace\_syntactic = [

    {'POS': 'VERB', 'ORTH': 'born'},

    {'OP': '\*'},

    {'LOWER': 'in', 'POS': 'ADP'},

    {'ENT\_TYPE': 'GPE', 'OP': '+'},

    {'IS\_PUNCT': True, 'OP': '\*'},

    {'ENT\_TYPE': 'GPE', 'OP': '\*'},

    {'IS\_PUNCT': True, 'OP': '\*'},

    ]

education\_syntactic = [

    {'TEXT': {'REGEX': '^(attend|attended|went|studied|University|College|trained|graduated)$'}},

    {'ENT\_TYPE': 'ORG', 'OP': '+'},

    ]

parents\_syntactic = [

    {'LOWER': {'REGEX': '^(born|son|daughter|parents)$'}},

    {'OP': '\*'},

    {'LOWER': {'REGEX': '^(to|of|are)$'}},

    {'TEXT': {'REGEX': '\s\*'}, 'OP': '\*'},

    {'IS\_PUNCT': True, 'OP': '\*'},

    {'ENT\_TYPE': 'PERSON', 'TEXT': {'REGEX': '\s\*'}, 'OP': '+'},

    {'IS\_PUNCT': True, 'OP': '\*'},

    {'LOWER': 'and','OP': '?'},

    {'TEXT': {'REGEX': '\s\*'}, 'OP': '\*'},

    {'IS\_PUNCT': True, 'OP': '\*'},

    {'ENT\_TYPE': 'PERSON', 'TEXT': {'REGEX': '\s\*'}, 'OP': '+'},

    {'IS\_PUNCT': True, 'OP': '\*'},

    ]

awards\_syntactic = [

    {'OP': '\*'},

    {'POS': 'NOUN', 'IS\_TITLE': '+', 'OP': '\*'},

    {'LOWER': {'REGEX': '^(winner|recipient|won|awarded|award-winning|nominated|nominee|nomination)$'}},

    {'POS': 'ADP', 'TEXT': {'REGEX': '^(of|the|for)$'}, 'OP': '\*'},

    {'POS': 'NOUN', 'IS\_TITLE': '+', 'OP': '\*'},

    {'OP': '\*'},

    ]

performances\_syntactic = [

    {'LOWER': {'REGEX': '^(star|starred|known|appeared|appearing|appearances|debute|album|recorded|play|played|role)$'}},

    {'POS': 'ADP', 'TEXT': {'REGEX': '^(in|of|the|for|as)$'}, 'OP': '+'},

    {'OP': '+'},

    ]

colleagues\_syntactic = [

    {'OP': '\*'},

    {'LOWER': {'REGEX': '^(starring|collaborating|working|worked|partnering|sharing.\*screen.\*space|appears.\*opposite|opposite|joined)$'}},

    {'LOWER': {'REGEX': '^(with|alongside|to)$'}}, {'OP': '\*'},

    {'ENT\_TYPE': 'PERSON'},

    {'OP': '\*'},

    ]

**Task 2.3:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Recall** | | **Precision** | |
|  | **Lexical** | **Syntactic** | **Lexical** | **Syntactic** |
| **Birthplace** | 68.4% | 68.4% | 75.0% | 73.3% |
| **Education** | 25.0% | 25.0% | 11.1% | 100% |
| **Parents** | 100% | 100% | 60.0% | 60.0% |
| **Awards** | 14.3% | 71.4% | 42.9% | 37.5% |
| **Performances** | 54.6% | 46.5% | 67.9% | 73.8% |
| **Colleagues** | 72.7% | 72.7% | 100% | 100% |