Author Disambiguation for Patents filed at USPTO

Swapnil Mishra
U5053816
Australian National University
Supervisor: Dr. Wray Buntine
Outline

• Background
• Patent Data Overview
• Approach and Goals
• Questions
Background

- **Disambiguation**
  - A sub field in Record Linkages
  - Matching individual records in a single database

- **USPTO Patents**
  - No unique and consistent identifier for inventors.
  - Inventors apply with different and partial information

- **Approaches**
  - String Matching with different linear weights on fields
  - Semi supervised classification
  - Supervised classification with help of labelled data
Patent Data Overview

*US6034431A*

Goosen et al.  (45) Date of Patent: Mar. 07, 2000

(64) Electronic integrated circuit with optical inputs and outputs
(75) Inventors: Keith Wayne Goosen, Aberdeen, New Jersey (US); Fouad E. Kiamilev, Charlotte, North Carolina (US); Ashok V. Krishnamoorthy, Middletown, New Jersey (US); David Andrew Barclay Miller, Fair Haven, New Jersey (US); and James Albert Walker, Howell, New Jersey (US)
(73) Assignees: Lucent Technologies, Inc., Murray Hill, New Jersey; Univ. of North Carolina, Charlotte

*US5605856A*

Goosen et al.  (45) Date of Patent: Feb. 25, 1997

(54) Method for designing an electronic integrated circuit with optical inputs and outputs
(75) Inventors: Keith W. Goosen, Aberdeen, New Jersey (US); Fouad E. Kiamilev, Charlotte, North Carolina (US); Ashok V. Krishnamoorthy, Middletown, New Jersey (US); David A. B. Miller, Fair Haven, New Jersey (US); and James A. Walker, Howell, New Jersey (US)
(73) Assignees: University of North Carolina, Charlotte, North Carolina; Lucent Technologies Inc., Murray Hill, New Jersey
Approach and Goals

- Learn existing algorithms and their usefulness.
- Build a semi supervised disambiguation engine to identify unique inventors.
- Find a way to use logistic regression with monotonic discrete features.
- Intelligent choices with help of external sources like patent families.
- A model that encapsulates effect of rare and common names in it.
- Evaluate against existing systems.
Questions?