A Linguistic Characterization of Google+ Posts across Different Social Groups
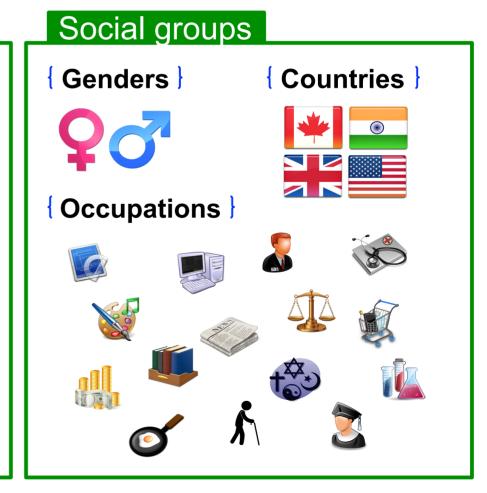
◆

Evandro Cunha, Gabriel Magno, Marcos André Gonçalves, César Cambraia, Virgilio Almeida

evandrocunha@dcc.ufmg.br, magno@dcc.ufmg.br

Universidade Federal de Minas Gerais, Brazil

Motivation

- Characterize G+ status updates
  - The textual genre "post"
  - Collective aspects of OSN users

- Detection of users' characteristics

- Applications
  - Fake entities in social groups
  - Personalization of services

Social groups

{ **Genders** }

{ **Countries** }

{ **Occupations** }

A Linguistic Characterization of Google+ Posts across Different Social Groups

Dataset

- About 2 million posts

Main findings

- Different groups =
    - Different structures
    - Different vocabulary

- Women → social and familial relations
- Men → technical topics and achievements

- Professional vocabulary
    - A professional OSN?

Analyses and results

- Fraction of misspellings per post
- Readability and structural complexity
- Semantic categories of words