

# A case for Keywords

- ▶ A large number of false matches had no reference to the relation involved.
- ▶ Eg. No mention of Population in the following matches:
  - ▶ The website of **China's** Ministry of Defense (MOD) has attracted around **1.25** billion visits in the three months since its opening, with the United States topping the source countries for foreign visits, website editor-in-chief Ji Guilin said.
  - ▶ Insulza, for his part, said the Organization of American States expects to raise **10 million** dollars for **Haiti's** recovery.
  - ▶ Koloini and others brought 10 million euros, probably **15 million**, back from **Iraq** at the time, Falter quoted from the diary.
- ▶ No reference to Co2 emission:
  - ▶ **China's** iron ore imports surged 41.6 percent to **627.8 million tonnes** in 2009, with the value falling 17.4 percent as prices were hit by the global downturn, customs data shows.

# A case for Keywords

## Good News

Sentences expressing a numerical relation can be expected to have keywords that denote the relation

- ▶ Take all the labeled sentences, prune out sentences that don't have atleast one of the relevant keywords

Internet User %	"Internet"
Land Area	"area", "land", "land area"
Population	"Population"
Diesel	"diesel"
GDP	"Gross domestic", "GDP"
CO2	"Carbon", "Carbon Emission", "CO2"
Inflation	"Inflation", "Price Rise"
FDI	"Foreign", "FDI"
Goods Export	"goods"
Life Expectancy	"life", "life expectancy"
Electricity Production	"Electricity"

# A case for Keywords

- ▶ Numbers are the second entity in our setup (Relation(Country, Number))
- ▶ Unlike real world entities, numbers don't have an identity of their own, sentences should have words (keywords!) indicating what the number stands for
- ▶ Manual inspection of 400 sentences pruned out after applying keyword based filter backs this conjecture, not even one false negative
- ▶ The keywords are created manually, can this process be automated?

Sentence

Extract Country  
Number Pairs

For each Country Number Pair

KB

Match Found for Relation R

Training  
Data

Add S with Label R

Process  
Next Sentence