

Numbers are incomplete without units

Numbers are incomplete without units

- Apart from being a constant quantity, a number usually makes sense when presented along with units.

Numbers are incomplete without units

- Apart from being a constant quantity, a number usually makes sense when presented along with units.

According to the [International Monetary Fund \(IMF\)](#), as of 2013, the Indian economy is nominally worth [US\\$1.842 trillion](#); it is the eleventh-largest economy by market exchange rates, and is, at [US\\$4.962 trillion](#), the third-largest by [purchasing power parity](#), or PPP.^[9] With its average annual GDP growth rate of [5.8%](#) over the past two decades, and reaching

Numbers are incomplete without units

- Apart from being a constant quantity, a number usually makes sense when presented along with units.

According to the [International Monetary Fund \(IMF\)](#), as of 2013, the Indian economy is nominally worth [US\\$1.842 trillion](#); it is the eleventh-largest economy by market exchange rates, and is, at [US\\$4.962 trillion](#), the third-largest by [purchasing power parity](#), or PPP.^[9] With its average annual GDP growth rate of [5.8%](#) over the past two decades, and reaching

- Units help in improving recall.

Numbers are incomplete without units

- Apart from being a constant quantity, a number usually makes sense when presented along with units.

According to the [International Monetary Fund](#) (IMF), as of 2013, the Indian economy is nominally worth ~~US\$1.842 trillion~~; it is the eleventh-largest economy by market exchange rates, and is, at ~~US\$4.962 trillion~~, the third-largest by [purchasing power parity](#), or PPP.^[9] With its average annual GDP growth rate of ~~5.8%~~ over the past two decades, and reaching

- Units help in improving recall.
 - In above example 1.842 as number alone would not match with the fact regarding economy in knowledge base.

Numbers are incomplete without units

- Apart from being a constant quantity, a number usually makes sense when presented along with units.

According to the [International Monetary Fund \(IMF\)](#), as of 2013, the Indian economy is nominally worth [US\\$1.842 trillion](#); it is the eleventh-largest economy by market exchange rates, and is, at [US\\$4.962 trillion](#), the third-largest by [purchasing power parity](#), or PPP.^[9] With its average annual GDP growth rate of [5.8%](#) over the past two decades, and reaching

- Units help in improving recall.
 - In above example 1.842 as number alone would not match with the fact regarding economy in knowledge base.
 - But with the unit trillion USD, we can normalize the value and then it would match existing facts in knowledge base.

Numbers are incomplete without units

According to the [International Monetary Fund](#) (IMF), as of 2013, the Indian economy is nominally worth **US\$1.842 trillion**; it is the eleventh-largest economy by market exchange rates, and is, at **US\$4.962 trillion**, the third-largest by [purchasing power parity](#), or PPP.^[9] With its average annual GDP growth rate of **5.8%** over the past two decades, and reaching

- Units help in reducing false positives and hence improving precision.

Numbers are incomplete without units

According to the [International Monetary Fund](#) (IMF), as of 2013, the Indian economy is nominally worth **US\$1.842 trillion**; it is the eleventh-largest economy by market exchange rates, and is, at **US\$4.962 trillion**, the third-largest by [purchasing power parity](#), or PPP.^[9] With its average annual GDP growth rate of **5.8%** over the past two decades, and reaching

- Units help in reducing false positives and hence improving precision.
 - If there is a fact, e.g, **inflation(India, 1.842%)** in knowledge base, then ignoring units can cause an incorrect match which leads in learning towards noisy patterns.

Unit Extraction is not easy!

Unit Extraction is not easy!

- Different ways to represent a single unit.

Unit Extraction is not easy!

- Different ways to represent a single unit.
 - Tunisia occupies an area of 163,610 **square kilometres**, of which 8,250 are water.

Unit Extraction is not easy!

- Different ways to represent a single unit.
 - Tunisia occupies an area of 163,610 **square kilometres**, of which 8,250 are water.
 - With an area of about 9.6 **million km^2** , the People's Republic of China is the 3rd largest country in total area behind Russia and Canada, and very similar to the United States.

Unit Extraction is not easy!

- Different ways to represent a single unit.
 - Tunisia occupies an area of 163,610 **square kilometres**, of which 8,250 are water.
 - With an area of about 9.6 **million km^2** , the People's Republic of China is the 3rd largest country in total area behind Russia and Canada, and very similar to the United States.
- Multiple units to represent a single numerical fact.

Unit Extraction is not easy!

- Different ways to represent a single unit.
 - Tunisia occupies an area of 163,610 **square kilometres**, of which 8,250 are water.
 - With an area of about 9.6 **million km^2** , the People's Republic of China is the 3rd largest country in total area behind Russia and Canada, and very similar to the United States.
- Multiple units to represent a single numerical fact.
 - Vatican City, a walled enclave within the city of Rome, with an area of approximately **44 hectares (110 acres)**, and a population of 842, is the smallest internationally recognized independent state in the world by both area and population.

Overview of Unit Extraction System

Overview of Unit Extraction System

- A discriminative context free grammar with scores attached to each possible production in the grammar.

Overview of Unit Extraction System

- A discriminative context free grammar with scores attached to each possible production in the grammar.
- A production P in the grammar is of the form $R ::= R_1 R_2$, scored as

$$\text{score}(P) = \mathbf{w} \cdot \mathbf{f}(P, x, i, j, k),$$

where (i, j) and $(j + 1, k)$ are text spans in x that R_1 and R_2 cover.

Overview of Unit Extraction System

- Some of the features that grammar uses to assign the best scores to various parses are as follows:

Overview of Unit Extraction System

- Some of the features that grammar uses to assign the best scores to various parses are as follows:
 - Matches with Unit Catalog

Overview of Unit Extraction System

- Some of the features that grammar uses to assign the best scores to various parses are as follows:
 - Matches with Unit Catalog
 - Lexical Clues

Overview of Unit Extraction System

- Some of the features that grammar uses to assign the best scores to various parses are as follows:
 - Matches with Unit Catalog
 - Lexical Clues
 - Relative Frequency - Prior of the word to be present as unit, then as a non-unit word. This is derived from WordNet ontologies.

Overview of Unit Extraction System

- Some of the features that grammar uses to assign the best scores to various parses are as follows:
 - Matches with Unit Catalog
 - Lexical Clues
 - Relative Frequency - Prior of the word to be present as unit, then as a non-unit word. This is derived from WordNet ontologies.
 - Co-occurrence statistics - presence of strongly co-occurring words in the text can help in disambiguating the various candidate units