

Semantic Search for Quantity Expressions

Tom Wiesing

Supervisor: Michael Kohlhase

Co-supervisor: Tobias Preusser

May 20, 2015

110392 Guided Research Applied and Computational
Mathematics & Thesis

Overview

Overview

- ▶ Motivation: Problem and State Of The Art

Overview

- ▶ Motivation: Problem and State Of The Art
- ▶ Our Approach: Structure Of The Search Engine

Overview

- ▶ Motivation: Problem and State Of The Art
- ▶ Our Approach: Structure Of The Search Engine
- ▶ The Implementation

Overview

- ▶ Motivation: Problem and State Of The Art
- ▶ Our Approach: Structure Of The Search Engine
- ▶ The Implementation
- ▶ Time for Questions

Motivation (1)

- ▶ We use units every day

Motivation (1)


- ▶ We use units every day
- ▶ We encounter them everywhere:

Motivation (1)


- ▶ We use units every day
- ▶ We encounter them everywhere:
 - ▶ When driving, there are speed limits, for example:




Motivation (1)

- ▶ We use units every day
- ▶ We encounter them everywhere:
 - ▶ When driving, there are speed limits, for example:  $\frac{\text{km}}{\text{h}}$
 - ▶ When baking, it often says in recipes something like: “add 3 tea spoons of sugar”


Motivation (1)

- ▶ We use units every day
- ▶ We encounter them everywhere:
 - ▶ When driving, there are speed limits, for example:  $\frac{\text{km}}{\text{h}}$
 - ▶ When baking, it often says in recipes something like: “add 3 tea spoons of sugar”
 - ▶ When shopping for shoes there are different sizes


Motivation (1)

- ▶ We use units every day
- ▶ We encounter them everywhere:
 - ▶ When driving, there are speed limits, for example:  $\frac{\text{km}}{\text{h}}$
 - ▶ When baking, it often says in recipes something like: “add 3 tea spoons of sugar”
 - ▶ When shopping for shoes there are different sizes
- ▶ In scientific papers they occur a lot

Motivation (1)

- ▶ We use units every day
- ▶ We encounter them everywhere:
 - ▶ When driving, there are speed limits, for example:  $\frac{\text{km}}{\text{h}}$
 - ▶ When baking, it often says in recipes something like: “add 3 tea spoons of sugar”
 - ▶ When shopping for shoes there are different sizes
- ▶ In scientific papers they occur a lot
- ▶ everything which somehow models a real system has at least one quantity expression

Motivation (1)

- ▶ We use units every day
- ▶ We encounter them everywhere:
 - ▶ When driving, there are speed limits, for example:  $\frac{\text{km}}{\text{h}}$
 - ▶ When baking, it often says in recipes something like: “add 3 tea spoons of sugar”
 - ▶ When shopping for shoes there are different sizes
- ▶ In scientific papers they occur a lot
- ▶ everything which somehow models a real system has at least one quantity expression
- ▶ everything is quantified

Motivation (2)

- ▶ But where is the problem?

Motivation (2)

- ▶ But where is the problem?
- ▶ Within one paper only a handful of units is used

Motivation (2)

- ▶ But where is the problem?
- ▶ Within one paper only a handful of units is used
- ▶ In general there are **a lot** of **different** units to describe **the same** quantity

Motivation (2)

- ▶ But where is the problem?
- ▶ Within one paper only a handful of units is used
- ▶ In general there are **a lot** of **different** units to describe **the same** quantity
- ▶ Just for lengths:

Motivation (2)

- ▶ But where is the problem?
- ▶ Within one paper only a handful of units is used
- ▶ In general there are **a lot** of **different** units to describe **the same** quantity
- ▶ Just for lengths:

Motivation (2)

- ▶ But where is the problem?
- ▶ Within one paper only a handful of units is used
- ▶ In general there are **a lot** of **different** units to describe **the same** quantity
- ▶ Just for lengths: *Meter*,

Motivation (2)

- ▶ But where is the problem?
- ▶ Within one paper only a handful of units is used
- ▶ In general there are **a lot** of **different** units to describe **the same** quantity
- ▶ Just for lengths: *Meter, Inch,*

Motivation (2)

- ▶ But where is the problem?
- ▶ Within one paper only a handful of units is used
- ▶ In general there are **a lot** of **different** units to describe **the same** quantity
- ▶ Just for lengths: *Meter, Inch, Foot,*

Motivation (2)

- ▶ But where is the problem?
- ▶ Within one paper only a handful of units is used
- ▶ In general there are **a lot** of **different** units to describe **the same** quantity
- ▶ Just for lengths: *Meter, Inch, Foot, Mile,*

Motivation (2)

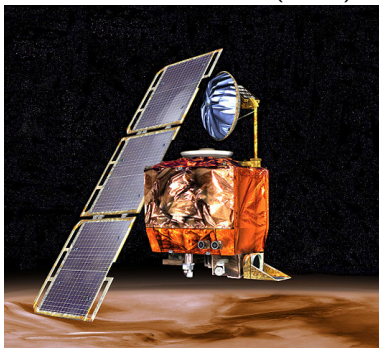
- ▶ But where is the problem?
- ▶ Within one paper only a handful of units is used
- ▶ In general there are **a lot** of **different** units to describe **the same** quantity
- ▶ Just for lengths: *Meter, Inch, Foot, Mile, Nautical Mile,*

Motivation (2)

- ▶ But where is the problem?
- ▶ Within one paper only a handful of units is used
- ▶ In general there are **a lot** of **different** units to describe **the same** quantity
- ▶ Just for lengths: *Meter, Inch, Foot, Mile, Nautical Mile, ...*
- ▶ This can cause problems when not converting properly

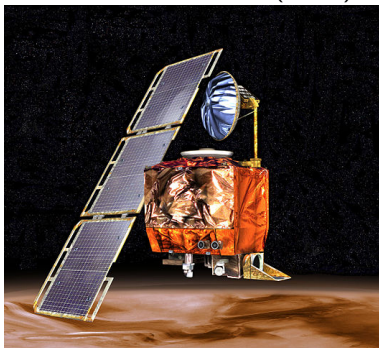
Motivation (2)

- ▶ But where is the problem?
- ▶ Within one paper only a handful of units is used
- ▶ In general there are **a lot** of **different** units to describe **the same** quantity
- ▶ Just for lengths: *Meter, Inch, Foot, Mile, Nautical Mile, ...*
- ▶ This can cause problems when not converting properly
- ▶ Mars Climate Orbiter (1999)



Motivation (2)

- ▶ But where is the problem?
- ▶ Within one paper only a handful of units is used
- ▶ In general there are **a lot** of **different** units to describe **the same** quantity
- ▶ Just for lengths: *Meter, Inch, Foot, Mile, Nautical Mile, ...*
- ▶ This can cause problems when not converting properly
- ▶ Mars Climate Orbiter (1999)



- ▶ Different Units are a big problem

Motivation (3)

Motivation (3)

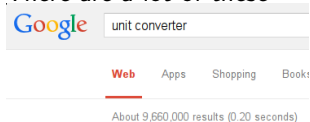
- ▶ What is the most common solution?

Motivation (3)

- ▶ What is the most common solution?
- ▶ Unit Converters

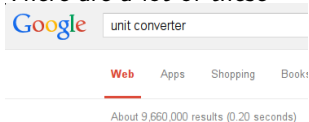
Motivation (3)

- ▶ What is the most common solution?
- ▶ Unit Converters
 - ▶ There are a lot of these

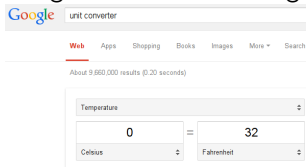


Motivation (3)

- ▶ What is the most common solution?
- ▶ Unit Converters
 - ▶ There are a lot of these



- ▶ Google itself has one integrated



Motivation (4)

- ▶ user needs to find out that conversion is required

Motivation (4)

- ▶ user needs to find out that conversion is required
- ▶ both input **and** output units need to be **given**

Motivation (4)

- ▶ user needs to find out that conversion is required
- ▶ both input **and** output units need to be **given**
- ▶ this is not integrated into the search process itself

Motivation (4)

- ▶ user needs to find out that conversion is required
- ▶ both input **and** output units need to be **given**
- ▶ this is not integrated into the search process itself
- ▶ Wouldn't it be nice:

Motivation (4)

- ▶ user needs to find out that conversion is required
- ▶ both input **and** output units need to be **given**
- ▶ this is not integrated into the search process itself
- ▶ Wouldn't it be nice:
 - ▶ when searching for $25 \frac{\text{m}}{\text{s}}$

Motivation (4)

- ▶ user needs to find out that conversion is required
- ▶ both input **and** output units need to be **given**
- ▶ this is not integrated into the search process itself
- ▶ Wouldn't it be nice:
 - ▶ when searching for $25 \frac{\text{m}}{\text{s}}$
 - ▶ we also find $90 \frac{\text{km}}{\text{h}}$

Motivation (4)

- ▶ user needs to find out that conversion is required
- ▶ both input **and** output units need to be **given**
- ▶ this is not integrated into the search process itself
- ▶ Wouldn't it be nice:
 - ▶ when searching for $25 \frac{\text{m}}{\text{s}}$
 - ▶ we also find $90 \frac{\text{km}}{\text{h}}$
 - ▶ we did not have to search for all the representations of the same quantity expression

Motivation (4)

- ▶ user needs to find out that conversion is required
- ▶ both input **and** output units need to be **given**
- ▶ this is not integrated into the search process itself
- ▶ Wouldn't it be nice:
 - ▶ when searching for $25 \frac{\text{m}}{\text{s}}$
 - ▶ we also find $90 \frac{\text{km}}{\text{h}}$
 - ▶ we did not have to search for all the representations of the same quantity expression
- ▶ This is the kind of search engine we have built

Our Approach (1)

- ▶ What components do we need for a semantic search engine?

Our Approach (1)

- ▶ What components do we need for a semantic search engine?
 1. A *Unit System* that is aware of the different representations of a QE

Our Approach (1)

- ▶ What components do we need for a semantic search engine?
 1. A *Unit System* that is aware of the different representations of a QE
 2. A *Spotter* that finds representations of QEs inside documents

Our Approach (1)

- ▶ What components do we need for a semantic search engine?
 1. A *Unit System* that is aware of the different representations of a QE
 2. A *Spotter* that finds representations of QEs inside documents
 3. A *Search Algorithm* that given a QE finds all its representations in the system

Our Approach (1)

- ▶ What components do we need for a semantic search engine?
 1. A *Unit System* that is aware of the different representations of a QE
 2. A *Spotter* that finds representations of QEs inside documents
 3. A *Search Algorithm* that given a QE finds all its representations in the system
 4. A *Frontend* that allows queries to be made

Our Approach (1)

- ▶ What components do we need for a semantic search engine?
 1. A *Unit System* that is aware of the different representations of a QE
 2. A *Spotter* that finds representations of QEs inside documents
 3. A *Search Algorithm* that given a QE finds all its representations in the system
 4. A *Frontend* that allows queries to be made
- ▶ Spotter is done by *Stiv Sherko*

Our Approach (1)

- ▶ What components do we need for a semantic search engine?
 1. A *Unit System* that is aware of the different representations of a QE
 2. A *Spotter* that finds representations of QEs inside documents
 3. A *Search Algorithm* that given a QE finds all its representations in the system
 4. A *Frontend* that allows queries to be made
- ▶ Spotter is done by *Stiv Sherko*
- ▶ We only need to worry about the *Unit System*, the *Search Algorithm* and the *Frontend*

Our Approach: The Unit System (1)

- ▶ Let's start with the Unit System

Our Approach: The Unit System (1)

- ▶ Let's start with the Unit System
- ▶ We orient ourselves on the SI specification

Our Approach: The Unit System (1)

- ▶ Let's start with the Unit System
- ▶ We orient ourselves on the SI specification
- ▶ Each quantity has a dimension

Our Approach: The Unit System (1)

- ▶ Let's start with the Unit System
- ▶ We orient ourselves on the SI specification
- ▶ Each quantity has a dimension
- ▶ According to SI there are 7 basic ones:

Our Approach: The Unit System (1)

- ▶ Let's start with the Unit System
- ▶ We orient ourselves on the SI specification
- ▶ Each quantity has a dimension
- ▶ According to SI there are 7 basic ones:
 - ▶ length

Our Approach: The Unit System (1)

- ▶ Let's start with the Unit System
- ▶ We orient ourselves on the SI specification
- ▶ Each quantity has a dimension
- ▶ According to SI there are 7 basic ones:
 - ▶ length
 - ▶ mass

Our Approach: The Unit System (1)

- ▶ Let's start with the Unit System
- ▶ We orient ourselves on the SI specification
- ▶ Each quantity has a dimension
- ▶ According to SI there are 7 basic ones:
 - ▶ length
 - ▶ mass
 - ▶ time

Our Approach: The Unit System (1)

- ▶ Let's start with the Unit System
- ▶ We orient ourselves on the SI specification
- ▶ Each quantity has a dimension
- ▶ According to SI there are 7 basic ones:
 - ▶ length
 - ▶ mass
 - ▶ time
 - ▶ electric current

Our Approach: The Unit System (1)

- ▶ Let's start with the Unit System
- ▶ We orient ourselves on the SI specification
- ▶ Each quantity has a dimension
- ▶ According to SI there are 7 basic ones:
 - ▶ length
 - ▶ mass
 - ▶ time
 - ▶ electric current
 - ▶ temperature

Our Approach: The Unit System (1)

- ▶ Let's start with the Unit System
- ▶ We orient ourselves on the SI specification
- ▶ Each quantity has a dimension
- ▶ According to SI there are 7 basic ones:
 - ▶ length
 - ▶ mass
 - ▶ time
 - ▶ electric current
 - ▶ temperature
 - ▶ luminous intensity

Our Approach: The Unit System (1)

- ▶ Let's start with the Unit System
- ▶ We orient ourselves on the SI specification
- ▶ Each quantity has a dimension
- ▶ According to SI there are 7 basic ones:
 - ▶ length
 - ▶ mass
 - ▶ time
 - ▶ electric current
 - ▶ temperature
 - ▶ luminous intensity
 - ▶ amount of substance

Our Approach: The Unit System (1)

- ▶ Let's start with the Unit System
- ▶ We orient ourselves on the SI specification
- ▶ Each quantity has a dimension
- ▶ According to SI there are 7 basic ones:
 - ▶ length
 - ▶ mass
 - ▶ time
 - ▶ electric current
 - ▶ temperature
 - ▶ luminous intensity
 - ▶ amount of substance
- ▶ but there are also quantities where we just *count*

Our Approach: The Unit System (1)

- ▶ Let's start with the Unit System
- ▶ We orient ourselves on the SI specification
- ▶ Each quantity has a dimension
- ▶ According to SI there are 7 basic ones:
 - ▶ length
 - ▶ mass
 - ▶ time
 - ▶ electric current
 - ▶ temperature
 - ▶ luminous intensity
 - ▶ amount of substance
- ▶ but there are also quantities where we just *count*
- ▶ and *dimensionless quantities* (such as Information)

Our Approach: The Unit System (1)

- ▶ Let's start with the Unit System
- ▶ We orient ourselves on the SI specification
- ▶ Each quantity has a dimension
- ▶ According to SI there are 7 basic ones:
 - ▶ length
 - ▶ mass
 - ▶ time
 - ▶ electric current
 - ▶ temperature
 - ▶ luminous intensity
 - ▶ amount of substance
- ▶ but there are also quantities where we just *count*
- ▶ and *dimensionless quantities* (such as Information)
- ▶ so we have 9 basic dimensions

Our Approach: The Unit System (2)

- ▶ we can also multiply these to get new dimensions

Our Approach: The Unit System (2)

- ▶ we can also multiply these to get new dimensions
- ▶ such as $\text{area} = \text{length} \cdot \text{length}$

Our Approach: The Unit System (2)

- ▶ we can also multiply these to get new dimensions
- ▶ such as $\text{area} = \text{length} \cdot \text{length}$
- ▶ similarly we can divide dimensions

Our Approach: The Unit System (2)

- ▶ we can also multiply these to get new dimensions
- ▶ such as $\text{area} = \text{length} \cdot \text{length}$
- ▶ similarly we can divide dimensions
- ▶ We can use this to get a meta-mathematical formalisation of dimensions

Our Approach: The Unit System (3)

Dimension		
dimension	:	type
none	:	dimension
count	:	dimension
length	:	dimension
mass	:	dimension
time	:	dimension
current	:	dimension
temperature	:	dimension
luminous	:	dimension
amount	:	dimension
.	:	dimension → dimension → dimension
\	:	dimension → dimension → dimension