Getting the Units Right

Moritz Schubotz & David Veenhuis & Howard S. Cohl

Outline

1. **Introduction**
2. Related Works
3. Proposed Approach
4. Conclusions / Future Works

Problem Statement

- Units important for understanding formulae
- Getting the dimensions right can be hard
- $a = bc^2 \qquad E = mc^2$
- This requires the knowledge that:
- E stands for energy, m for mass and c for speed of light and hence possible units are [J] = [Nm] = [kgm²/s²]
- Once the dimensions are know the units are straight forward to calculate
- Easy for humans, hard for computers

Wikidata Basics

- Free knowledge base for structured data
- Machine readable form
- Triple storage
  - example: ("**energy (Q11379)**", "instance of", "scalar physical quantity"), ("**energy (Q11379)**", "instance of", "extensive physical property")
- Access to data via API
- Expandable data model

WIKIDATA

Outline

1. Introduction
2. **Related Works**
3. Proposed Approach
4. Conclusions / Future Works

Related Works

- Extraction of candidates for identifier using natural language processing [Schubotz16]
- Dimensional analysis to find consistent unit assignments for the identifier [Liew02]
- Mathematical type for physical variables [Collins08]
- Support for mathematical formulae in Wikidata [Hilbig16]

Outline

1. Introduction
2. Related Works
3. **Proposed Approach**
4. Conclusions / Future Works

Proposed Approach

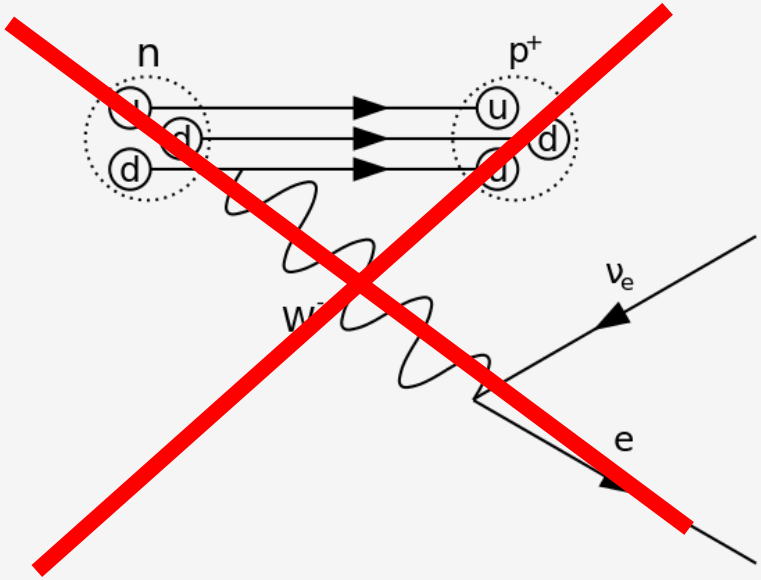
1. Identify physical formulae
   1. Find formulae in Wikipedia
   2. Detect identifiers in formulae
   3. Decide if identifiers are physical quantities by finding corresponding dimensions /units from Wikidata
2. Automatically derive units of involved identifiers
3. Update Wikidata to include learned data

Restrictions

- <u>Definition</u> A **physical formula** is a *binary* mathematical relation of type *equation* or *inequality* containing one or more physical quantities.
- Only basic mathematical operations
- Focus on K12 formulae

$$\boldsymbol{W} = \int_{\vec{s}_1}^{\vec{s}_2} \vec{\boldsymbol{F}}(\vec{\boldsymbol{s}}) \cdot \mathbf{d}\vec{\boldsymbol{s}}$$

$$\boldsymbol{F} = \frac{\mathbf{1}}{\mathbf{4}\boldsymbol{\pi}\boldsymbol{\varepsilon}_{\mathbf{0}}} \frac{\boldsymbol{q}_{\mathbf{1}} \cdot \boldsymbol{q}_{\mathbf{2}}}{\boldsymbol{r}^{\mathbf{2}}}$$



Unit Inference

- <u>Definition</u>: A physical formula is **valid** if the dimensions on both sides are compatible.

class	**rule**	**constraint**	**operators**
3: map	$o\left(\begin{bmatrix}\mathsf{a}\\ x\end{bmatrix}, \begin{bmatrix}\mathsf{b}\\ y\end{bmatrix}\right) \rightarrow \begin{bmatrix}\mathsf{o(a,b)}\\ \frac{o(x,y)}{\lvert o(x,y)\rvert_1}\end{bmatrix}$		times, division, integration, differentiation
2: restrict	$o\left(\begin{bmatrix}\mathsf{a}\\ x\end{bmatrix}, \begin{bmatrix}\mathsf{b}\\ y\end{bmatrix}\right) \rightarrow \begin{bmatrix}\mathsf{o(a,b)}\\ x\end{bmatrix}$	$x = y$	plus, minus, equals
1: apply	$o\left(\begin{bmatrix}\mathsf{a}\\ x\end{bmatrix}, \begin{bmatrix}\mathsf{b}\\ y\end{bmatrix}\right) \rightarrow \begin{bmatrix}\mathsf{o(a,b)}\\ o(x,b)\end{bmatrix}$	$y = 1$	power, roots
0: unitless	$o\left(\begin{bmatrix}\mathsf{a}\\ y\end{bmatrix}\right) \rightarrow \begin{bmatrix}\mathsf{o(a)}\\ y\end{bmatrix}$	$y = 1$	function application

Unit Inference

- Example formula Mass-energy equivalence

$$E = mc^2$$

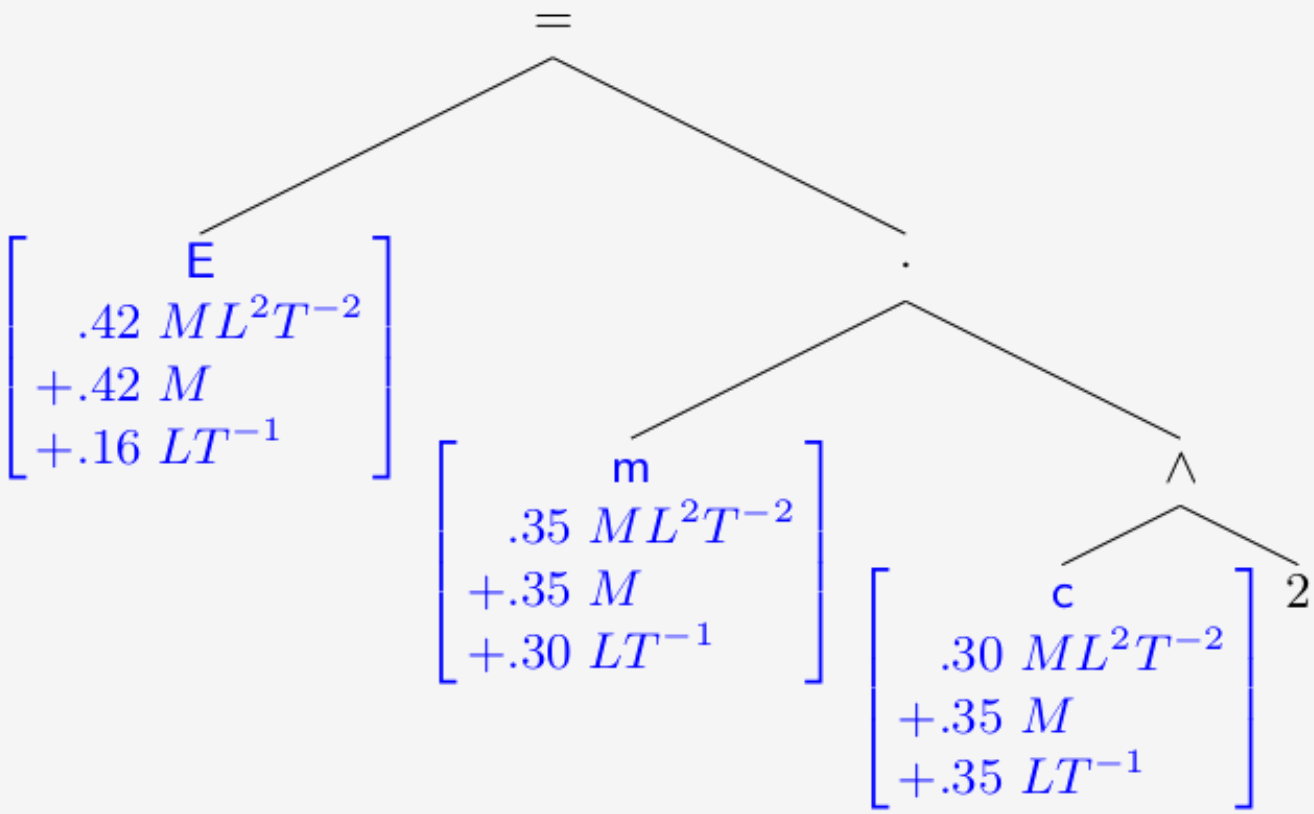
- Tree algorithm to infer the dimensions

Id	definition	score
E	energy	0.42
E	mass	0.42
E	speed of light	0.16

Id	definition	score
m	energy	0.35
m	mass	0.35
m	speed of light	0.30

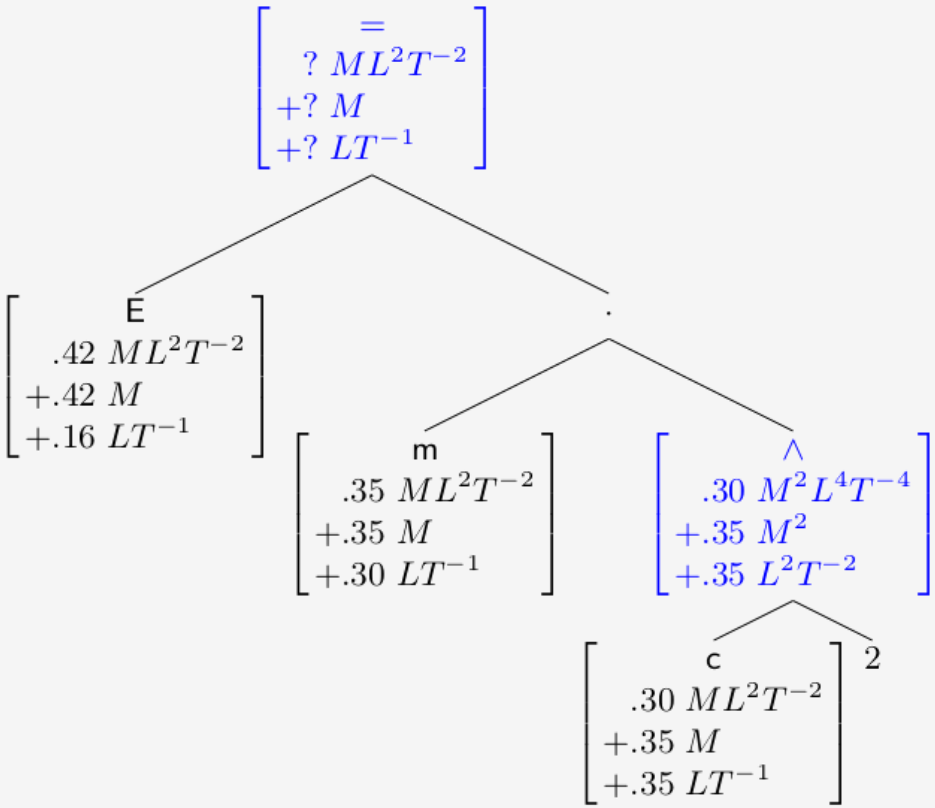
Id	definition	score
c	energy	0.30
c	mass	0.35
c	speed of light	0.35

Unit Inference: Algorithm



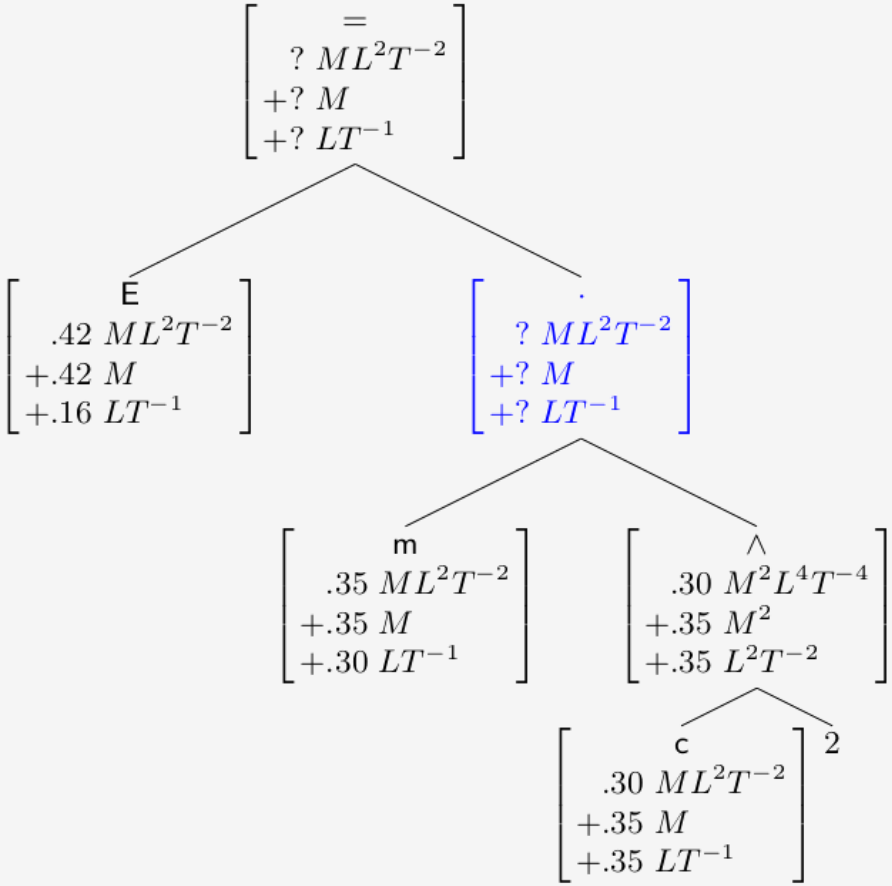
(1) Set identifier dimensions

Unit Inference: Algorithm



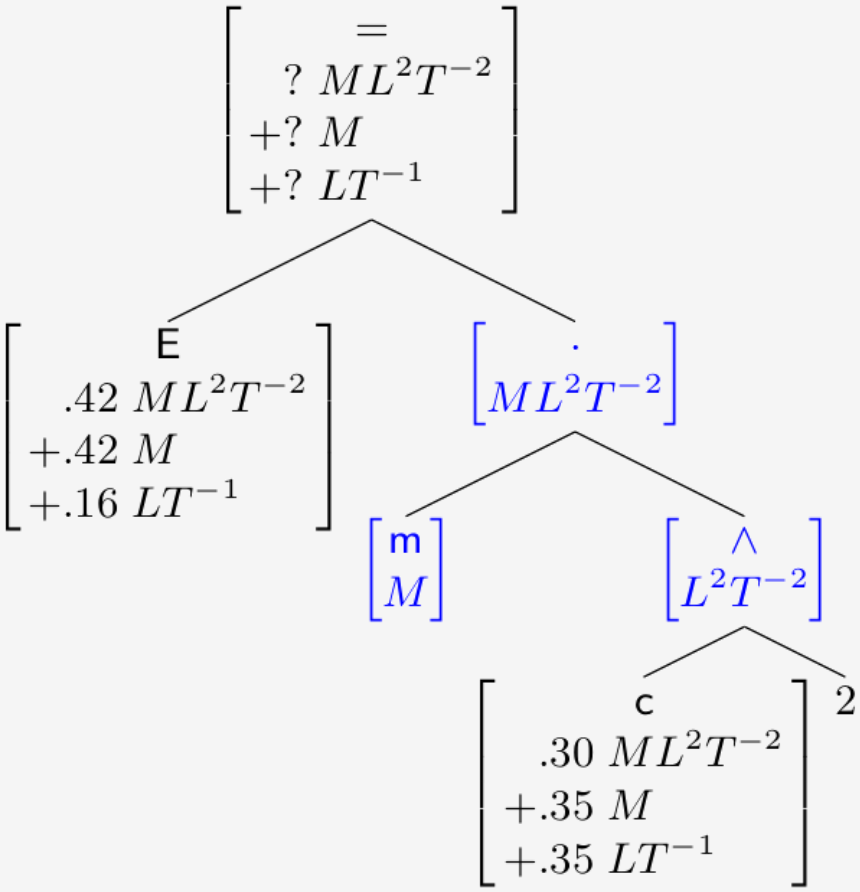
(2) Propagate upwards

Unit Inference: Algorithm



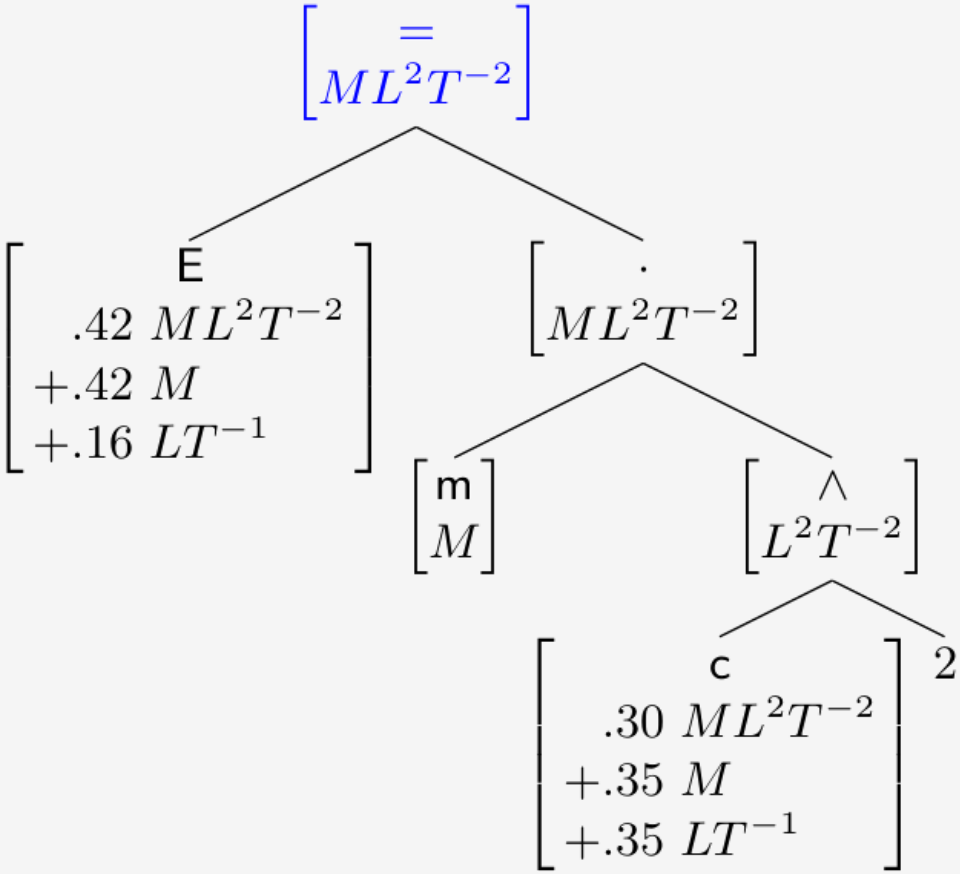
(3) Propagate downwards

Unit Inference: Algorithm



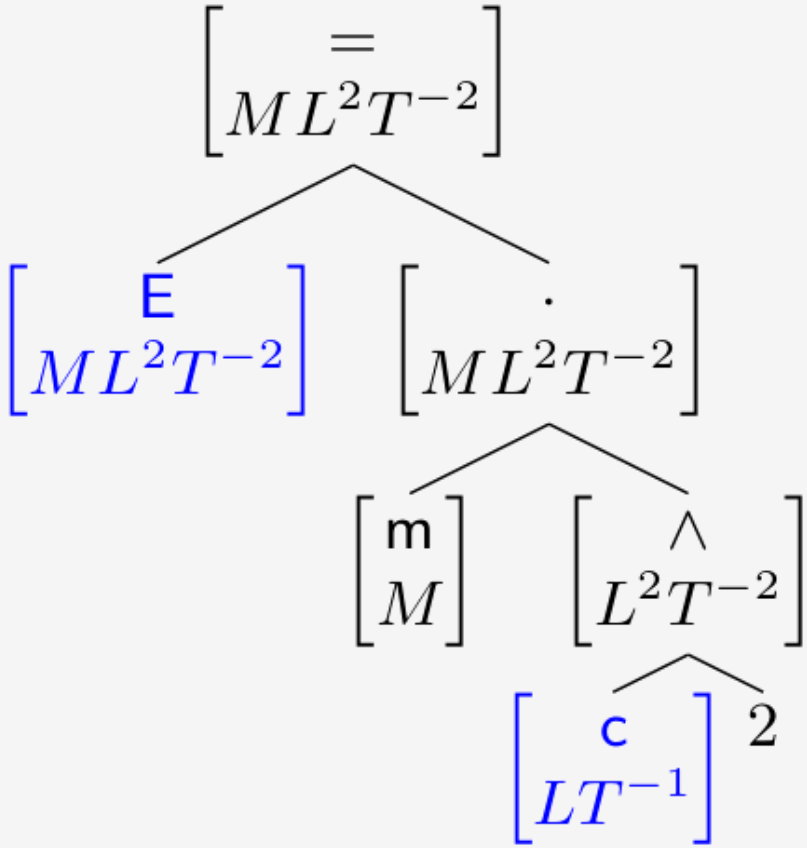
(4) Process maps

Unit Inference: Algorithm



(5) Propagate upwards

Unit Inference: Algorithm



(6) Propagate downwards

Updating Wikidata

- unique solution for identifier-definition mapping?
- Items for formula already exists?
- If not create item
- Link every identifier to appropriate item of its definition in Wikidata. Following the links the units can be found

Outline

1. Introduction
2. Related Works
3. Proposed Approach
4. **Conclusions / Future Works**

Conclusions and Future Works

- Automatically updating formulae is possible
- Depends on quality of existing Wikidata items
- Automatic import and export to Wikidata needs to be done
- Improvement by stemming since definitions result from NLP

Contact

Moritz Schubotz (now at Universität Konstanz)

moritz@schubotz.de

+49 7531 88 4438

Mobile: +49 1578 047 1397

www.isg.uni-konstanz.de

www.formulasearchengine.com