

# GROBID quantities

Quantities measurement are a set of conventions for which specific symbols are used to express concepts and informations...

*...unfortunately there are still too many different systems (Metric, Imperial, etc), and too much freedom.*

# With natural language, the ambiguity is increasing

Fifty-three journals were collected: 13 were eliminated from analysis [...]

The cells (  $1 \times 10^7$  ) were incubated in RPMI-1640 medium containing [...]

# Input is mostly noisy



I am typing this on a manual typewriter. It's not very easy but ~~it~~ at least doesn't mean I am using electricity pointlessly. I bought the typewriter on eBay, juts as I earlier bought an anñient treadle-powered sewing machine...

\\LT.HE; Eifos  
'SEETHE  
OF OUFIGOD"ISaluuu.  
THE OUTLOOK.  
AN EARNEST WORD.  
N 1880 we had but 35 Missionaries, all told, in South America.  
Now we have 76, or more than double the number that we  
had 19 years ago. It would naturally be supposed that the  
income of the S.A.M.S. would have increased in something  
like the same proportion. Has this been the case? No!  
On the contrary, the income from all sources remains almost  
i stationary. For instance, the average income for the past 18  
years has been ,£12,305, while the actual income for the now distant year  
iSSo was ,£11,836, or only .£669 less than tint average. The income for  
181/1 was but .£12,361, or only ^525 more thr.n in 1880 ' In 1X97 it was  
£'7i573, but that was most exceptional, as it included one noble donation  
of ,£2,600.  
To state the present position in another way. the work of the Society  
for Cod and South America has increased 117 per cent., while the money  
entrusted to the Society, with which to sustain that work, has increased  
very little. It is manifest that such a result has not been attained  
APRIL, 1891,1

● They use 20 g of proteins to prepare the solution.

Language Identifier

EN They use 20g of proteins to prepare the solution.

Raw unstructured data

Quantity parser  
CRF

EN They use 20 g of proteins to prepare the solution.

Raw structured data

Value parser  
CRF  
Lexicon

Unit parser  
CRF  
Lexicon

value  
20 20.0  
number

unit  
g 0.001 kg  
mass

Parsed data

Quantity normaliser  
JSR-363

2000000.0 g → 2000 kg

Normalised data

● They use 0.002 g of proteins to prepare the solution.

substance  
mass

Demo