Aligning Measurement Units Across Varied Data Sources

Mihran Simonian - 12386294
mihran.simonian@gmail.com
University of Amsterdam

ABSTRACT

While the manufacturing industry attempts to implement open data communication over the whole supply chain, existing computer (ERP) systems are maintained in-place at individual manufacturing plants. The wide range of systems, together with the wide range of sensors used to measure manufacturing parameters, create multiple challenges. As supply chains operate and procure more cross-boarder than ever before, there is a growing need for a standardized units of measurements. However, computer systems are hardly designed to cope with measurement units, let alone when introducing multiple units. In this paper I propose a system to unify these units and will go into details of issues that might arise when executing the proposed procedures on a larger cluster of computers.

1 INTRODUCTION

The manufacturing industry is currently going through a major development, described as 'Smart Industry 4.0'. By using computers, big data and the implementation of various sensors, manufacturers attempt to optimize their production flows and lower costs, whilst increasing reliability and optimize production speeds.

Simultaneously OEM producers (manufacturers whom combine loose parts to assemble one product) attempt to optimize their supply chain, by combining the data streams of various suppliers attempt to optimize it's supply chain, in order to avoid production halts, which are very costly and can create significant delivery delays to the end-users.

Due to the fact that production facilities can possibly be spread all over the world[1], there is an increasing issue that is rising but currently less highlighted as computer systems are currently not necessarily designed to cope with multiple measurement units[2].

Differences in measurement units used in manufacturing industry are visible on all type of fields, yet temperatures, forces, distances, volume and weight are the main types (we can derive most others from these types). To illustrate the magnitude of the issue, in 1999 NASA lost one of it's satellites orbiting the planet Mars, due to a programming error in which an Imperial unit was implemented but the Metric unit had to be used instead[2] [1]. Surely manufacturing a car or assembling parts for a lithography machine[3] will not result in the same financial (and social) consequences, but it is a sign that we should not discard the issue as it will grow larger if we don't tackle it now.

2 RELATED WORK

The issue of using multiple sensors and vastly storing these in computer systems has already been highlighted in previous work [3], in which data variable types were highlighted as possible sources of conflict. This was in 1990, however in 2013 there was still reason enough for concern, as [2] described in his paper.

Co-currently the Dutch research institute TNO is currently developing (test version live since january 2020) a communication protocol in cooperation with various manufacturers[4][5]. This communication standard enables manufacturers to share various production parameters (data) in an easy manner, whilst the manufactures remain to be able to use their own ERP[6] system, due to various reasons (legacy purposes, proprietary software, vendor lock-ins, etc.). The reason of existence for these proposed systems is to assure that companies do not need to 'give up' their own dedicated ERP systems, but at the same time be enabled to share data with other manufacturers (cross ERP). Thus, systems might communicate via live data connections or by means of backlogs, where on a daily night backlogs are loaded into systems (by using csv file containers for instance). The main issue currently faced by these organizations, is scaling up the system and determining the level of information sharing for an optimum improvement.

3 RESEARCH

During this research I would like to answer the following question:

- How do we assure a synchronous setting of measurement units as used in a typical manufacturing plant?

[1] https://www.asml.com/en/company/sustainability/responsible-supply-chain

[2] https://mars.jpl.nasa.gov/msp98/news/mco991110.html

[3] https://www.asml.com/en/technology

[4] https://smartindustry.nl/fieldlabs/8-smart-connected-supplier-network/

[5] https://www.brainportindustries.com/nl/berichten/maakindustrie-aan-de-slag-met-digitalisering

[6] $https://en.wikipedia.org/wiki/Enterprise_resource_planning$

With sub questions:

- Which type of data variables do we want to align in the current scope?
- Which possible inputs can we expect?
- Which problems might arise during the scale-up of the system?

In case this will not grow to large, I would also like to discuss this bonus question:

- How should we pre-clean the data entries?

4 METHOD

By automating the processing of data entries, we can minimize the human efforts required (and all issues that might arise from this). In order to automate we need to set up a 'translation list' of conversions between Imperial (US) and Metric (SI) units. Furthermore we need to assure that users provide an input for the translation program to understand which unit is used, in order to translate it to the correct conversion.

By storing all the inputs into a standard (metric by preference) system, we are assured that all information as stored into this system follows one standard unit system. Output conversions must than occur in case this is needed.

Furthermore I would like to discuss how the scaling up of this system would occur and which issues might arise. For instance, what parts of this data process can we perform on a cluster of computers and which not?

Lastly I would also like to test whether more procedures can be written to assure that data input is separated accordingly (string conversions, data variable type conversions). This might lead to an extensive exercise, which in itself can be a complete paper so this one I would like to keep for the last part and see more as a bonus section if there is sufficient time left.

To summarize:

- Write a short but sufficient dictionary in python for unit conversion
- Write multiple functions using Python (exact libraries unknown, probably only Pandas)
- Test system
- Discuss potential issues upon implementation on clusters

Using:

- CSV
- XML
- SQL
- Python and Pandas

ACKNOWLEDGMENTS

To Hannes Mühlheisen, for the fun and joy during the lectures and for setting up this excercise, together with Cristian Rodriguez Rivero.

For Shuo Chen; My apologies for this lengthy description, but I hope you will appreciate the content!

REFERENCES

[1] MCO Mishap Investigation Board. 1999. *Mars Climate Orbiter Mishap Investigation Board Phase.* https://llis.nasa.gov/llis_lib/pdf/1009464main1_0641-mr.pdf

[2] Marcus P Foster. 2013. Quantities, units and computing. *Computer Standards & Interfaces* 35, 5 (2013), 529–535.

[3] Edward Waltz, James Llinas, et al. 1990. *Multisensor data fusion.* Vol. 685. Artech house Boston.