

# **Magellan: Toward Building Entity Matching Management Systems**

This paper presents new Entity Matching System called Magellan. The authors mention that most research on EM focused on algorithms not on EM systems. Also, current EM systems lack of entire EM pipeline support and guides. Magellan tries to solve these problems. It has step-by-step how-to guides, external Python tools support and it is an open-world system. Paper refers each of these features and the authors challenges to develop Magellan system.

Magellan is a huge system thus authors mentions only a subset of the whole system in this paper. That is a good thing because if they were describe each of the system component in this paper, it would be very long system documentation report. Open-world system for EM idea is very clever and good for users. With that, users could choose from variety of Python data tools and use what they want. Also, guides for each step is very useful for the user that are not familiar with either data management, data domain or EM algorithms.

The paper doesn't contain too much technical information. It only presents new system and its some components. There are few examples and some of them use lots of letters. Too much letters may cause confusion on readers. Also, they could use table graph examples in order to improve examples effect. In Selecting a Matcher section, there is no description about how to choose features or how did they implement it.

More examples with visuals would be great. Also, they could have mentioned the open-world systems vs closed-world system part a little bit longer for better understanding. I don't know if other Magellan papers refers to Python stacks but this paper doesn't have any information about how to use external Python tools and scripting within Magellan system.

2019-01-27

Baran Kaya, [kayab@mcmaster.ca](mailto:kayab@mcmaster.ca), 400284996