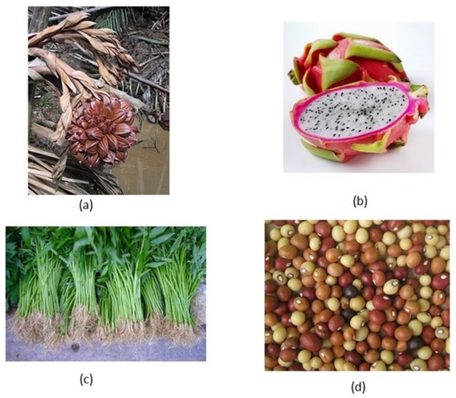
Quality Assessment Methods for Crowd-sourced Data of Underutilised Crops

Chi Dai Huynh
Supervisors: Prof. Mike Jackson, Dr. Tuong Thuy Vu, and Dr. Didier Leibovici

What are underutilised crops?

Underutilised crops refer to crop species whose potential is often neglected. However, they have many promising characteristics for improving food security, e.g. helping subsistence and income of the poor, reducing risk of over-relying on limited major crops, increasing agricultural sustainability, contributing to food quality, and maintaining the diversity of cultural and dietary (Mayes et al., 2012).



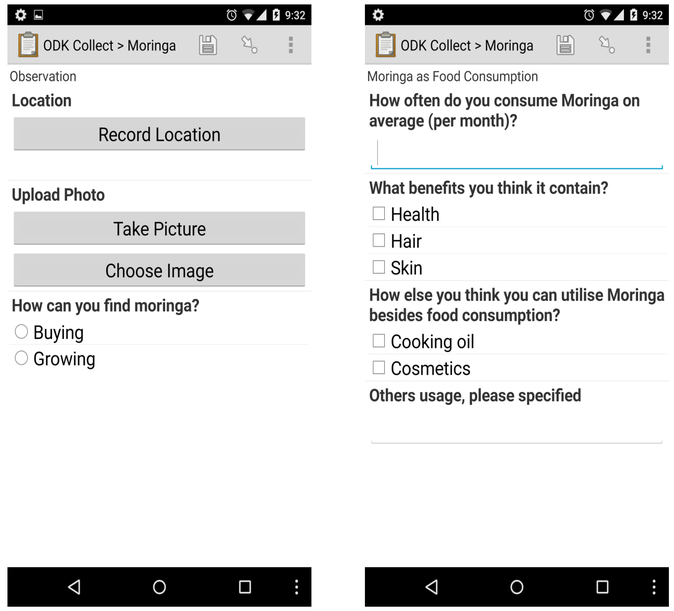
Examples of Underutilised Crops: (a) Nipa palm, (b) Dragon fruit, (c) Water spinach, and (d) Bambara groundnut.

Research Motivation

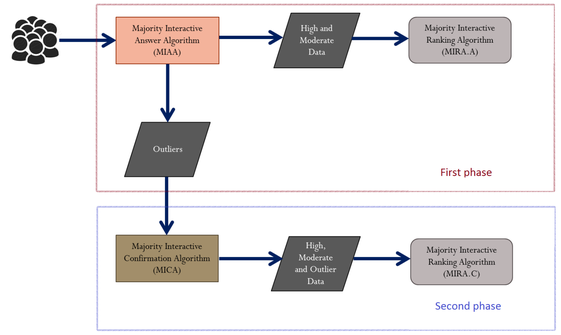
There is a lack of data to use for new research on underutilised crops. Collecting authoritative data needs a data collection infrastructure and is expensive, especially as underutilised crops are typically grown in small and dispersed farms. Crowdsourcing offers a possible means of overcoming this data deficiency at low cost. However, unlike traditional authoritative data, crowdsourcing is not collected using QA processes (Idris et al., 2014). Consequently, this raises our interests in finding a range of methods and algorithmic techniques that can validate, parameterize, and improve the quality of submitted crowd-sourced underutilised crops data.

Experiments

A study of benefits and usage (e.g., food consumption) of moringa will be conducted to measure the knowledge of local people who consume and/or grow moringa. This study also examines their knowledge and ability of recognizing moringa among other underutilised crops. A survey containing a list of photos and queries regarding moringa and its usage will be distributed to crowd-volunteers who are asked for their responses. Finally, a generic volunteer-centric quality assessment approach is proposed, and applied to validate the quality of crowdsourced volunteer responses.



Methodology



A Generic Volunteer-centric Approach for Assuring the Quality of Crowd-sourced Data (Huynh et al., 2014)

Conclusion

The research is focused on deriving techniques to quality assure crowd-sourced data so that it may be used for scientific studies and predictive modeling. The research will develop methods and algorithms that may be used within the “seven-pillars” standards-based QA framework developed at University of Nottingham (Meek et al., 2014), and where there is little authoritative data available to assist the QA process. The use case selected is the crowdsourcing of data for research into underutilised crops.

References

HUYNH, D., VU, T., JACKSON, M., LEIBOVICI, D. & MAYES, S. (2014) *Towards a User-Centric Approach for Assessing Quality of Crowd-sourced Data, an Application in Agriculture*. Positional paper for GIScience 2014 Workshop on Role of Volunteered Geographic Information in Advancing Science: Effective Utilization.

IDRIS, N. H., JACKSON, M. J. & ISHAK, M. H. I. (2014) *A conceptual model of the automated credibility assessment of the volunteered geographic information*. IOP Conference Series: Earth and Environmental Science, 18.

MAYES, S., MASSAWE, F. J., ALDERSON, P. G., ROBERTS, J. A., AZAM-ALI, S. N. & HERMANN, M. (2012) *The potential for underutilised crops to improve security of food production*. Journal of Experimental Botany, 63, 1075 - 1079.

MEEK, S., JACKSON, M. & LEIBOVICI, D. (2014) *A flexible framework for assessing the quality of crowd-sourced data*. AGILE conference, 3-6 June 2014, Castellón, Spain.