Information Reputation

Peter Davis[1] and Salman Haq[2]

[1] Distinguished Engineer,
Neustar, 21575 Ridgetop Circle, Sterling, Virginia, USA
`peter.davis@neustar.biz`,
[2] Research Architect,
Neustar, 21575 Ridgetop Circle, Sterling, VA, USA
`salman.haq@neustar.biz`

Abstract. In this paper we describe the design of a *reputation framework* for an information management system under active development. The integration of a reputation framework with an IMS is a novel combination that can produce a distinctly more effective business intelligence tool.

1 Introduction

Neustar is a data analytics and intelligence services company that operates several large database systems. To efficiently manage these numerous, disparate systems, we are developing an Information Management System (IMS) that maps technical data models using a standard set of ontologies. The IMS is an online community for employees where they can share, classify and discover metadata about various Neustar *data source*s. Its main purpose is to assist users in achieving two main objectives: *a)* reducing costs by utilizing existing information and *b)* increasing revenues by creating new information [3].

With these objectives in mind, users must have the ability to make value judgments about *data source*s relative to one another. Such *data source*s may number in the hundreds and the *datum* contained therein may number in the tens of thousands. The majority of these entities will lack significant value for data science, and those that are valuable will risk being lost in a deluge of information. Therefore it is imperative that the system establish a bias towards meaningful *datum* by highlighting *interestingness*. A well-crafted reputation framework can excel at doing exactly this.

2 Terminology

Glossary

claim One or more assertions made of a *datum*.
data source A computer system that stores data such as a database or file system.

data steward A individual or group of individuals holding domain-specific knowledge of an information system.

datum An instance of metadata mapped to an atomic data field. This includes, for example, columns in a relational database or entities defined in an XML schema.

interestingness A scalar value indicating the suitability for inclusion in further analysis.

reputation A qualitative measure that informs a value judgment about a *datum* or user.

Acronyms

IMS Information Management System.

3 Framework Description

The framework is comprised of several *reputation* models, each of which computes one or more scores for a resource type. A fixed set of *claim*s serve as inputs to each model which assigns numerical values to them and passes them through a series of mathematical filter *processes*. Models are distinguished by their input selection, process configuration, and output scores. The IMS utilizes a fixed ontology to define *claim*s that include appropriate business and technical classifications for data within the subject systems. The essential *claim*s of the *datum* model are summarized in Table 1. The IMS also incorporates techniques to simplify crowd sourcing the classification of *datum* by *data steward*s. However, we have concluded from early usage, that a simple classification process is insufficient. Classifications can be subjective, and classification sparseness results in under utilization of the system. As a consequence, methods to encourage accurate and complete classification will be implemented to enrich the overall efficacy of the system.

Table 1. Essential *Claim*s for the *Datum* reputation model

Name	Description
classified	*Data steward* classified *datum* from the business domain ontology
described	*Data steward* entered a description
discussed	User participated in a discussion topic about *datum*
emailed	User emailed the link to the *datum* page to another user
flagged	User informed the *data steward* about insufficient or inaccurate details
watched	User will be notified of future updates by other users
wanted	User requested access to the *datum* from the *data steward*

4 Interestingness Reputation Model for Datum

In the IMS, *datum* is an atomic unit of data. Its' classification results in queriable metadata, and can relate to a column in a relational database or an element, attribute or phrase in a document. At the time of writing, the system had *over 7,000 fields from merely four* data source*s*. Even at this watermark, the task of finding *interesting datum* is impractical for any user community. As more *data source*s are imported into the system, this task will be become impossible even if the datum population grows sub-linearly. Therefore it is imperative that the system is capable of identifying and highlighting interesting datum to facilitate user objectives.

In Figure 1 we describe the simplified model for calculating the *interestingness* reputation score for *datum*. Our approach is informed by [1] which applies a similar methodology for surfacing interesting media objects. The score is an indicator of the likelihood that a particular datum has potential value. User interactions with the IMS are interpreted as *claim*s from Table 1. The figure shows *claim*s as they are consumed by various processes. The intermediate processes (boxes 7, 8, 9, and 10) compute normalized counts of the *claim* interactions. These counts are fed to the terminal process, *InterestingnessCustomMixer* (box 12) which scales and reduces the values into the scalar *interestingness* score. This score can be used as a predictor for search and recommendation systems. Omitted from this simplified model are lag and decay filters necessary to counteract volatility and freshness bias respectively [2].

5 Conclusion And Future Work

We have described a realistic blue print for a reputation system that is on the roadmap of our IMS. Once implemented we think that it will dramatically improve the quality of information that is retrievable by users, thus increasing its' effectiveness as a platform for information management and data science. We have left outcome analysis of the approach and results for a future paper. Also on the roadmap is a meaningful gamification system inspired by [4] to further enhance user engagement.

References

1. Butterfield, Daniel, et al. Interestingness Ranking of Media Objects. Yahoo! Inc., assignee. Patent US20060242139A1. 8 Feb. 2006. Print.
2. Farmer, F. Randall., and Bryce Glass. Building Web Reputation Systems. Sebastopol, CA: O'Reilly, 2010. Print.
3. Sveiby, K.E. Knowledge Management: Lessons from the Pioneers. Tech. Sveiby Knowledge Associates, 2001. Web. 12 Mar. 2013. http://www.sveiby.com/articles/KM-lessons.doc
4. Nicholson, Scott. Strategies for Meaningful Gamification: Concepts behind Transformative Play and Participatory Museums. Tech. Meaningful Play 2012 Web. 13 Mar. 2013. http://scottnicholson.com/pubs/meaningfulstrategies.pdf
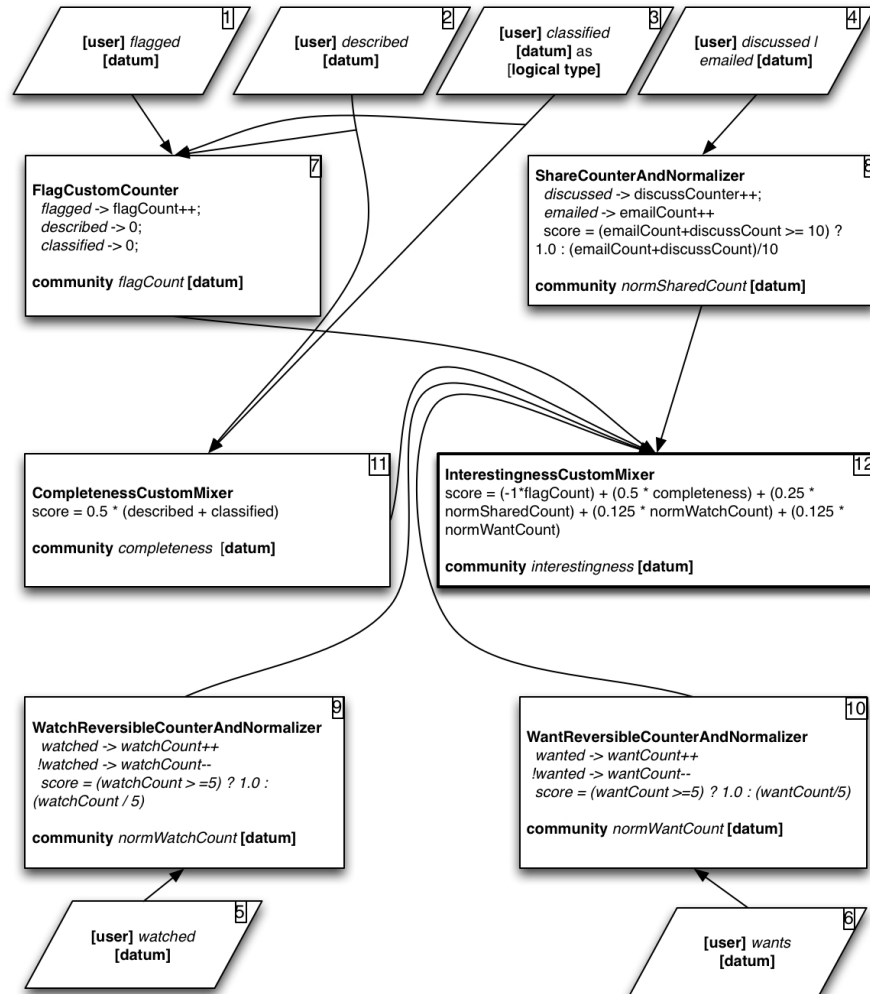
4

Fig. 1. Interestingness Reputation Model for Datum