Modeling and Reasoning about Information Quality Requirements

Mohamad Gharib and Paolo Giorgini

University of Trento - DISI,
38123, Povo, Trento, Italy

gharib, paolo.giorgini@disi.unitn.it

March 18, 2015

Outline

Introduction

Motivation Information Quality (IQ) requirements has to be considered since the early phases of software development.

Problem Most of the Requirements Engineering frameworks either ignore or loosely define IQ requirements.

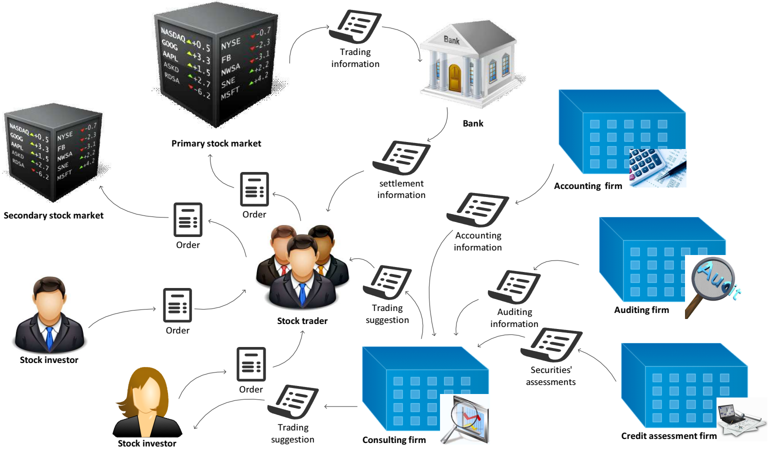
Principal ideas We propose a novel conceptual framework for modeling and reasoning about IQ at requirements level.

Contribution The utility of our proposed framework have been demonstrated by modeling and analyzing the IQ requirements of a case study concerning a main U.S stock market crash (the Flash Crash).

Problem statement

- Quality has been defined as “fitness for use” [6], or the conformance to specifications [14].
- IQ is a hierarchical multi-dimensional concept that can be characterized by several dimensions [9], and several models for analyzing IQ based on diffrent dimensions have been propose [20, 9, 13].
- Traditionally, IQ has been considered as a technical problem focusing mainly on how information are stored and transmitted by technical components (both hardware and software).
- Most of the existing IQ approaches and models do not satisfy the needs of current complex systems (e.g., socio-technical systems [4]).

Case Study: a stock market system



- **Actors**: *investors*, *traders*, *stock markets*, *consulting firms*, etc.
- **Assets**: *orders*, *consulting suggestions*, *CB information*, etc.

Multi-dimensional IQ model

Our framework considers 4 IQ dimensions to analyze IQ, namely: accuracy, completeness, timeliness, and consistency.

Accuracy: information should be true or error free with respect to some known, designated or measured value [2].

Multi-dimensional IQ model

Our framework considers 4 IQ dimensions to analyze IQ, namely: accuracy, completeness, timeliness, and consistency.

Accuracy: information should be true or error free with respect to some known, designated or measured value [2].

Example

Fraud information (falsified) that have been used by some actors played a role in the Flash Crash [7, 11].

- HFTs' flickering quotes that are orders last very short time, which make them unavailable for most of traders [11]).
- Market Makers' stub quotes that are orders with prices far away from the current market prices, such orders can also be considered as falsified information; since they are orders were not intended to be performed [7].

Multi-dimensional IQ model

Our framework consider 4 IQ dimensions to analyze IQ, namely: accuracy, completeness, timeliness, and consistency.

Accuracy: information should be true or error free with respect to some known, designated or measured value [2].

Completeness: all parts of information should be available [2, 18].

Example

The highly fragmented nature of the finical market along with the inefficient coordination mechanisms among the CBs of the trading markets also played a role in the Flash Crash [5, 17].

- If trading markets fail to coordinate their CBs, HFTs will simply search for a market other than the closed ones and continue trading [10]. For instance, during the Flash Crash CME employs its CB but NYSE did not [17].

Multi-dimensional IQ model

Our framework consider 4 IQ dimensions to analyze IQ, namely: accuracy, completeness, timeliness, and consistency.

- **Accuracy:** information should be true or error free with respect to some known, designated or measured value [2].
- **Completeness:** all parts of information should be available [2, 18].
- **Timeliness:** to which extent information is valid in terms of time [13].

Example

Nanex report [1] listed both NYSE-CQS[a] related information delay along with the DOW Jones delay (was a result of the first delay) as a reason of the Flash Crash.

[a] Consolidated Quotation System (CQS) information concerns quotation, where a quote is an order that has not been performed

Multi-dimensional IQ model

Our framework consider 4 IQ dimensions to analyze IQ, namely: accuracy, completeness, timeliness, and consistency.

Accuracy: information should be true or error free with respect to some known, designated or measured value [2].

Completeness: all parts of information should be available [2, 18].

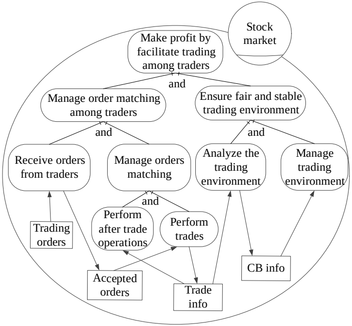
Timeliness: to which extent information is valid in terms of time [13].

Consistency: means that multiple recordings of the same information should be the same across time and space [2].

Example

To ensure efficient coordination among the CBs of the trading markets, we need to guarantee that these markets depend on consistent information for employing their CBs. According to [8], provision time from *CME* to *Nasdaq* was 13 (ms), while provision time from *CME* to *NYSE* was 14.65 (ms), which leads to inconsistent information between these two markets.

Extending secure Tropos with IQ modeling concepts

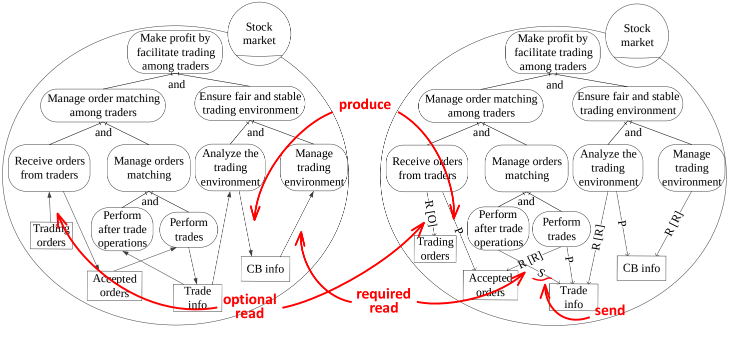


(a) secure Tropos

Our modeling language extends the secure Tropos modeling languages:

Goal-information relations: produces, optional/ required read, and sends.

Extending secure Tropos with IQ modeling concepts



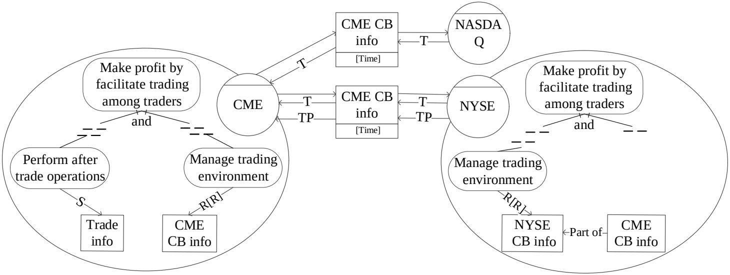
(a) secure Tropos

(b) extended secure Tropos

Our modeling language extends the secure Tropos modeling languages:

Goal-information relations: produces, optional/ required read, and sends.

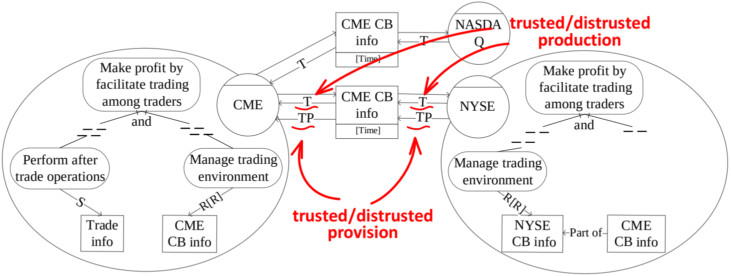
Extending secure Tropos with IQ modeling concepts



Our modeling language extends the secure Tropos modeling languages:

Information accuracy: we provide trusted/distrusted production, and trusted/distrusted provision concepts.

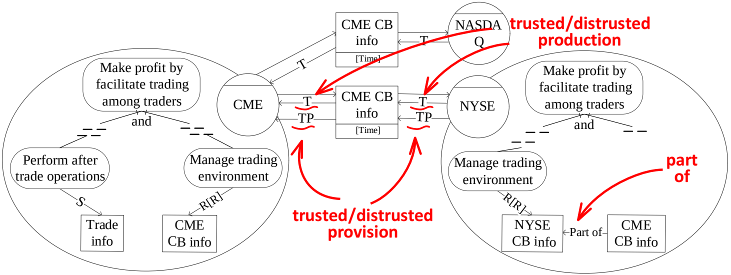
Extending secure Tropos with IQ modeling concepts



Our modeling language extends the secure Tropos modeling languages:

Information accuracy: we provide trusted/distrusted production, and trusted/distrusted provision concepts.

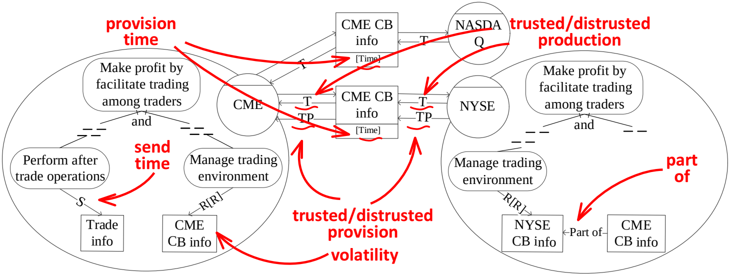
Extending secure Tropos with IQ modeling concepts



Our modeling language extends the secure Tropos modeling languages:

Information completeness: we provide the part of concept that enables for addressing information completeness.

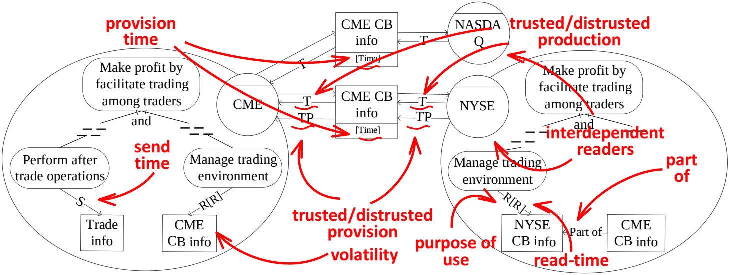
Extending secure Tropos with IQ modeling concepts



Our modeling language extends the secure Tropos modeling languages:

Information timeliness: we provide information volatility, read timeliness, send timeliness, and information provision time concepts.

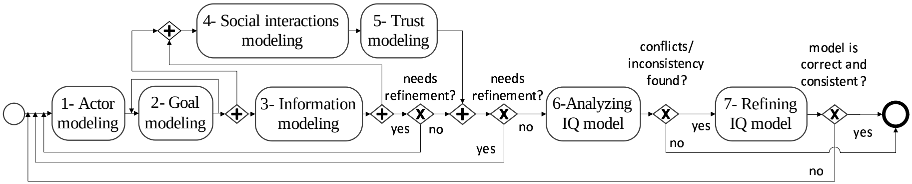
Extending secure Tropos with IQ modeling concepts



Our modeling language extends the secure Tropos modeling languages:

Information consistency: we provide read-time and interdependent readers concepts.

The methodological process



Our Framework is supported with an engineering methodology that is used for the systematic design of the system-to-be.

Modeling phase: aims to model IQ requirements of the system-to-be in their social and organizational context.

Analysis phase: aims to analyze the IQ requirements model against the properties of the design.

Refinement phase if some IQ requirements did not hold, the analysis tries to find proper solution for them at this phase.

Evaluation using the Flash Crash Scenario

We have demonstrated the utility of our proposed framework by modeling and analyzing IQ requirements of case study concerning a main U.S stock market crash (the Flash Crash).

Applicability and effectiveness : we developed a prototype implementation of our framework[1] to test its applicability and effectiveness for modeling and reasoning about IQ requirements.

Scalability of the reasoning technique: we investigate the reasoning execution time against the requirements model size, by increasing the number of its modeling elements from 188 to 1316 through 7 steps, and calculate the reasoning execution time at each step. The result shows that the relation between the size of the model and execution time is not exponential.

[1] http://mohamadgharib.wordpress.com/

Related Work

Improving IQ by design Wand and Wang [18], Total Data Quality Management (TDQM) methodology [19], IP-MAP [16], IP-UML [15]. However, all the previously mentioned approaches were not designed to capture neither the organizational nor the social aspects of the system-to-be, which are very important aspects in current complex systems.

Requirements Engineering RE community did not appropriately support modeling nor analyzing IQ related issues (e.g., KAOS [3], *i** [21], secure Tropos [12], etc.). Usually, they care about information in terms of its availability and who is responsible of providing it.

Conclusions and future work

Contribution: We proposed a framework that enables system designers for modeling and reasoning about IQ requirements from the early system design taking into consideration the intended purpose of information usage.
Future work:

- We intend to extend the IQ dimensions we considered, and investigate the interrelations among IQ dimensions.
- Refine some IQ related concepts.
- Information production process needs more investigation.

Nanex flash crash summary report.
http://www.nanex.net/FlashCrashFinal/FlashCrashSummary.html, 2010.
Accessed: 2014-05-30.

Bovee, M., Srivastava, R. P., and Mak, B.
A conceptual framework and belief-function approach to assessing overall information quality.
International journal of intelligent systems 18, 1 (2003), 51–74.

Dardenne, A., Van Lamsweerde, A., and Fickas, S.
Goal-directed requirements acquisition.
Science of computer programming 20, 1-2 (1993), 3–50.

Emery, F., and Trist, E.
Socio-technical systems. management sciences, models and techniques. churchman cw et al, 1960.

Gomber, P., Haferkorn, M., Lutat, M., and Zimmermann, K.
The effect of single-stock circuit breakers on the quality of fragmented markets.
In *FinanceCom* (2012), pp. 71–87.

Juran, J., Gryna, F., and Bingham, R.
Quality control handbook.
New York [etc.]: McGraw-Hill, 1979.

References II

KIRILENKO, A., KYLE, A. S., SAMADI, M., AND TUZUN, T.
The flash crash: The impact of high frequency trading on an electronic market.
Manuscript, U of Maryland (2011).

LEWIS, M.
Flash boys: a Wall Street revolt.
WW Norton & Company, 2014.

LIU, L., AND CHI, L.
Evolutional data quality: A theory-specific view.
In *IQ* (2002), pp. 292–304.

MADHAVAN, A.
Exchange-traded funds, market structure and the flash crash.
SSRN Electronic Journal (2011), 1–33.

MCINISH, T., AND UPSON, J.
Strategic liquidity supply in a market with fast and slow traders.
Available at SSRN 1924991 (2012).

MOURATIDIS, H., AND GIORGINI, P.
Secure tropos: A security-oriented extension of the tropos methodology.
International Journal of Software Engineering and Knowledge Engineering 17, 2 (2007), 285–309.

References III

Pipino, L. L., Lee, Y. W., and Wang, R. Y.
Data quality assessment.
Communications of the ACM 45, 4 (2002), 211–218.

Reeves, C. A., and Bednar, D. A.
Defining quality: alternatives and implications.
Academy of management Review 19, 3 (1994), 419–445.

Scannapieco, M., Pernici, B., and Pierce, E.
Ip-uml: Towards a methodology for quality improvement based on the ip-map framework.
In *7th Intl Conf. on Information Quality (ICIQ-02)* (2002), pp. 8–10.

Shankaranarayanan, G., Wang, R., and Ziad, M.
Ip-map: Representing the manufacture of an information product.
In *Proceedings of the 2000 Conference on Information Quality* (2000), pp. 1–16.

Subrahmanyam, A.
Algorithmic trading, the flash crash, and coordinated circuit breakers.
Borsa Istanbul Review 13, 3 (2013), 4–9.

Wand, Y., and Wang, R.
Anchoring data quality dimensions in ontological foundations.
Communications of the ACM 39, 11 (1996), 86–95.

References IV

WANG, R.
A product perspective on total data quality management.
Communications of the ACM 41, 2 (1998), 58–65.

WANG, R., AND STRONG, D.
Beyond accuracy: What data quality means to data consumers.
Journal of management information systems (1996), 5–33.

YU, E. S.-K.
Modelling strategic relationships for process reengineering.
PhD thesis, University of Toronto, 1995.

?

Question