INFOCOMP / DataSys 2019 International Expert Panel:

Panel on Information Processing

Preserving Quality during Information/Data Creation and Processing

July 30, 2019, Nice, France

The Ninth Int. Conference on Advanced Communications and Computation (INFOCOMP 2019)

INFOCOMP / DataSys
July 28–August 1, 2019 - Nice, France

INFOCOMP Expert Panel: . . . Preserving Quality . . .

Panelists

- *Claus-Peter Rückemann* (Moderator),
  Westfälische Wilhelms-Universität Münster (WWU) / KiM, DIMF / Leibniz Universität Hannover, Germany
- *Małgorzata Pankowska,*
  Department of Informatics, University of Economics in Katowice, Poland
- *Ahmed Gomaa,*
  University of Scranton, USA
- *Mauro Dell'Orco,*
  Polytechnic University of Bari, Italy

INFOCOMP 2019: `http://www.iaria.org/conferences2019/INFOCOMP19.html`
Program: `http://www.iaria.org/conferences2019/ProgramINFOCOMP19.html`
ISSN: 2308-3484, ISBN: 978-1-61208-732-0

INFOCOMP Expert Panel: . . . Preserving Quality . . .

Pre-Discussion-Wrapup/Statements: (Rückemann, Pankowska, Gomaa, Dell'Orco)

- 'Quality' of **knowledge, information, and data** creation and -development are essential for processing and preserving 'quality'.
- **Understanding facets and complementary attributes** of knowledge, information, and data as well as differences is essential for **formalisation** and management.
- 'Measures for quality' are **beyond mathematical/statistical-only-measures**.
- **Methodological approaches, systematics, and standardisation** can contribute to 'quality'.
- **Quality is a secondary virtue**.
- **Combining Design with Research** (Hevner approach, Design Science Research).
- Applicability of **Design Science Research in Data Governance**; Focus on Quality in the whole **Data Governance Process**.
- Does information quality mean precision or significance? Are there **different views**?
- What is the association between multimodal data processing and **meaningful results**?
- How far can **unstructured data** be used for creating actionable insights?
- Are there **'alternatives' to 'quality'**?

INFOCOMP Expert Panel: Panel Discussion Comments

Panel discussion, additional comments (Pankowska):

- Value of information: profits received when the decision is made with the information in comparison with profits when the decision is made without information
- Usability as quality feature for information as well as for software evaluation. Usability means that information can be accessible and we can use it for purposes we have established.
- Software quality is an abstract concept. Its presence is difficult to define, because of its complexity, However, the first and the most important characteristics of software quality is software usability. It means that software functionalities are accessible and used by users. They prefer just these functionalities. Software Quality is determined by fulfilment functional and non-functional requirements. The second covers reliability, efficiency, operability and security requirements. Beyond quality characteristics important for now, there are maintainability characteristics which determine abilities for future development, and transferability characteristics which determine ability to apply that piece of software on another platform, e.g., adaptability, installability, compliance.

INFOCOMP Expert Panel: Panel Discussion Comments

Panel discussion, additional comments (Rückemann):

- Always understand your ‘knowledge’. (improving non-technical, epistomological education)
- ‘Quality’ of knowledge, information, and data creation and data development are essential for processing and preserving ‘quality’.
- Understanding facets and complementary attributes of knowledge, information, and data as well as differences is essential for formalisation and management.
- Creating and preserving quality common ‘knowledge, information, and data’ processing scenarios require a detailed understanding of formal attributation.
- Always remember and practice (!) the principle of / expertise in formalisation.
- There are many scenarios, which cannot be dealt with using simple quality criteria, e.g., mathematical criteria.
- Start with quality of data creation and quality of input.
- Stick with keeping scenarios’ implementations strictly as what they are.
- Do not over-interpret ‘quality’, limit yourself to your scenario.
- Quality can be result of hermeticism.
- Consider the achievements of solid information science and best practice with knowledge, information, and data and associated value. Value of ’knowledge, information, data’ is ranked primary on the scale of values.

INFOCOMP Expert Panel: Panel Discussion Comments

Panel discussion add. comments; Best Practice & Def. (KiM, DIMF):

- *Rückemann, Claus-Peter; Pavani, Raffaella; Schubert, Lutz; Gersbeck-Schierholz, Birgit; Hülsmann, Friedrich; Lau, Olaf; and Hofmeister, Martin (2018): Post-Summit Results, Delegates' Summit: Best Practice and Definitions of Data Value; Sept. 13, 2018, The Eighth Symposium on Advanced Computation and Information in Natural and Applied Sciences (SACINAS), The 16th Internat. Conf. of Numerical Analysis and Applied Mathematics (ICNAAM), Sept. 13–18, 2018, Rhodos, Greece.*
  URL: `http://www.user.uni-hannover.de/cpr/x/publ/2018/delegatessummit2018/rueckemann_icnaam2018_summit_summary.pdf`,
  URL: `https://doi.org/10.15488/3639` *(DOI).*
- *Rückemann, Claus-Peter; Iakushkin, Oleg O.; Gersbeck-Schierholz Birgit; Hülsmann, Friedrich; Schubert, Lutz; and Lau, Olaf (2017): Post-Summit Results, Delegates' Summit: Best Practice and Definitions of Data Sciences – Beyond Statistics; Sept. 25, 2017, The Seventh Symposium on Advanced Computation and Information in Natural and Applied Sciences (SACINAS), The 15th Internat. Conf. of Numerical Analysis and Applied Mathematics (ICNAAM), Sept. 25–30, 2017, Thessaloniki, Greece.*
  URL: `http://www.user.uni-hannover.de/cpr/x/publ/2017/delegatessummit2017/rueckemann_icnaam2017_summit_summary.pdf`,
  URL: `https://www.tib.eu/en/search/id/datacite%3Adoi~10.15488%252F3411/Best-Practice-and-Definitions-of-Data-Sciences/`,
  URL: `https://doi.org/10.15488/3411` *(DOI).*
- *Rückemann, Claus-Peter; Kovacheva, Zlatinka; Schubert, Lutz; Lishchuk, Iryna; Gersbeck-Schierholz, Birgit; and Hülsmann, Friedrich (2016): Post-Summit Results, Delegates' Summit: Best Practice and Definitions of Data-centric and Big Data – Science, Society, Law, Industry, and Engineering; Sept. 19, 2016, The Sixth Symposium on Advanced Computation and Information in Natural and Applied Sciences (SACINAS), The 14th Internat. Conf. of Numerical Analysis and Applied Mathematics (ICNAAM), Sept. 19–25, 2016, Rhodes, Greece.*
  URL: `http://www.user.uni-hannover.de/cpr/x/publ/2016/delegatessummit2016/rueckemann_icnaam2016_summit_summary.pdf`,
  URL: `https://www.tib.eu/en/search/id/datacite%3Adoi~10.15488%252F3410/Best-Practice-and-Definitions-of-Data-centric-and/`,
  URL: `https://doi.org/10.15488/3410` *(DOI).*
- *Rückemann, Claus-Peter; Hülsmann, Friedrich; Gersbeck-Schierholz, Birgit; Skurowski, Przemyslaw; and Staniszewski, Michal (2015) Post-Summit Results, Delegates' Summit: Best Practice and Definitions of Knowledge and Computing; Sept. 23, 2015, The Fifth Symposium on Advanced Computation and Information in Natural and Applied Sciences (SACINAS), The 13th Internat. Conf. of Numerical Analysis and Applied Mathematics (ICNAAM), Sept. 23–29, 2015, Rhodes, Greece.*
  URL: `http://www.user.uni-hannover.de/cpr/x/publ/2015/delegatessummit2015/rueckemann_icnaam2015_summit_summary.pdf`,
  URL: `https://www.tib.eu/en/search/id/datacite%3Adoi~10.15488%252F3409/Best-Practice-and-Defnitions-of-Knowledge-and-Computing/`,
  URL: `https://doi.org/10.15488/3409` *(DOI).*

INFOCOMP Expert Panel: Post-Panel-Discussion Summary

Post-Panel-Discussion Summary (2019-08-10): (Rückemann, Pankowska, Gomaa, Dell'Orco)

- **Quality is not a property of a physical or non-physical entity or knowledge.**
- **Quality is a secondary virtue, depending on perspectives, defined criteria, targets, and goals.**
- **Quality/qualities depend(s) on the respective views, including, e.g., relativity, precision, verification, significance, usability, reliability, extendability, adaptability, securability, transferability, and perception. This is true for all targets, e.g., data, software, and services.**
- **Always be aware of hermeticism when postulating an "ultimate/absolute" view of quality.**
- **Knowledge is in itself "multi-dimensional" (information science). Knowledge can be considered to be associated with complements (e.g., after Aristotle and after Krathwohl/Anderson), e.g., factual, conceptual, procedural, and metacognitive knowledge. Complements can represent a respective view. Any such complements cannot be isolated in general.**
- **Intrinsic properties of knowledge can generally not be represented or transformed into extrinsic properties. Especially, knowledge itself cannot be 'stored'.**
- **There are arbitrary ways to represent information in its knowledge context. Remember, a picture can be worth a thousand words – and – a word can be worth a thousand pictures.**
- **Many scenarios can be formalised to some extent, generally including abstraction and reduction, which allows to apply 'tools'. Tools, e.g., any computer systems, are restricted to a certain formalisation, as, e.g., are classifications, ontologies, fuzzy logics, and neural networks. Implementations of tools and formalisations can allow to associate properties and quantities with the targetted approaches to quality.**

INFOCOMP Expert Panel: Table of Presentations, Attached

Panelist Presentations: (presentation sort order, following pages)

- **The Practical Significance of Information Science and Information Life-cycle** (*Rückemann*)
- **Design Science Research on Data Governance** (*Pankowska*)
- **Creating summarized actionable insights from tweets and generating meaningful composite multimedia objects: Lessons learned** (*Gomaa*)
- **Precision vs Significance: Which is the Meaning of Information Quality?** (*Dell'Orco*)

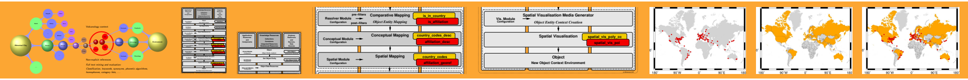
INFOCOMP / DataSys 2019 International Expert Panel:
Preserving Quality during Information/Data Creation and Processing

The Practical Significance of Information Science and Information Life-cycle

The Ninth Int. Conference on Advanced Communications and Computation (INFOCOMP 2019)
July 30, 2019, Nice, France

Dr. rer. nat. Claus-Peter Rückemann[1,2,3]

[1] Westfälische Wilhelms-Universität Münster (WWU), Münster, Germany
[2] KiM, DIMF, Germany
[3] Leibniz Universität Hannover, Hannover, Germany
ruckema(at)uni-muenster.de



Status of common perception and practice

- Common perception: *Quality is based on 'physical' property.*
- Common perception: *Quality can (always) be measured.*
- Common perception: *Quality is a result of processing, algorithms etc.*
- Common practice: *Quality criteria are made independent of input.*
- Common practice: *Often terms like 'premium' are used as 'alternatives' to 'quality' for various reasons.*

Status, facts commonly overseen

- **Quality is a secondary virtue**.
- **Fundaments of 'quality' and its measures are arbitrary**.
- Often, **quality approaches and practice result from hermeticism**.
- 'Measures for quality' are **beyond mathematical/statistical-only measures**.
- **Information Science, methodological approaches, systematics, and standardisation** can contribute to information/data 'quality'.

Vision

- Broad awareness that 'Quality' is a holistic undertaking.
- Broad awareness of principles, reasons, and consequences of hermeticism.
- 'Quality' endeavours need to understand information sciences' fundaments. (knowledge, information, data; knowledge complements and life-cycles; ...)
- Processing itself results from formalisation.
  'Quality' is a secondary, hermeticism approach.
  Any hermeticism approacher has to level with formalisation.
- Aware: Being of quality is beyond mathematical measures.
- Dependencies: 'Quality' of any output produced by any technique depends upon the intrinsic 'quality' of the input.
  (after many experiences and researchers, classical by Mayne and others).
- Consider: Content-context-structures; data-centric, systematical, methodological approaches; data-locality alternatives; implementation/realisation and associated processing aspects.
- Recognise that values and valorisation are beyond mathematical (economic) measures. Infrastructures and services can be secondary interest!
- Check methods and logic, e.g., reasoning for grid density; if considering mathematical 'quality' base, have error calculation for all algorithms and steps; check plausibility of each step ...

Conclusions: Quality is a result of hermeticism

- **Always understand your 'knowledge'.** (improving non-technical, epistomological education)
- 'Quality' of **knowledge, information, and data** creation and data development are essential for processing and preserving 'quality'.
- **Understanding facets and complementary attributes** of knowledge, information, and data as well as differences is essential for **formalisation** and management.
- Creating and preserving quality common 'knowledge, information, and data' processing scenarios require a detailed understanding of formal attributation.
- **Always remember and practice (!) the principle of / expertise in formalisation.**
- **There are many scenarios, which cannot be dealt with using simple quality criteria, e.g. mathematical criteria.**
- **Start with quality of data creation and quality of input.**
- **Stick with keeping scenarios' implementations strictly as what they are.**
- **Do not over-interpret 'quality', limit yourself to your scenario.**

University of Economics in Katowice

IARIA

your place

your space

your future

Preserving Quality during Information/Data Creation and Processing

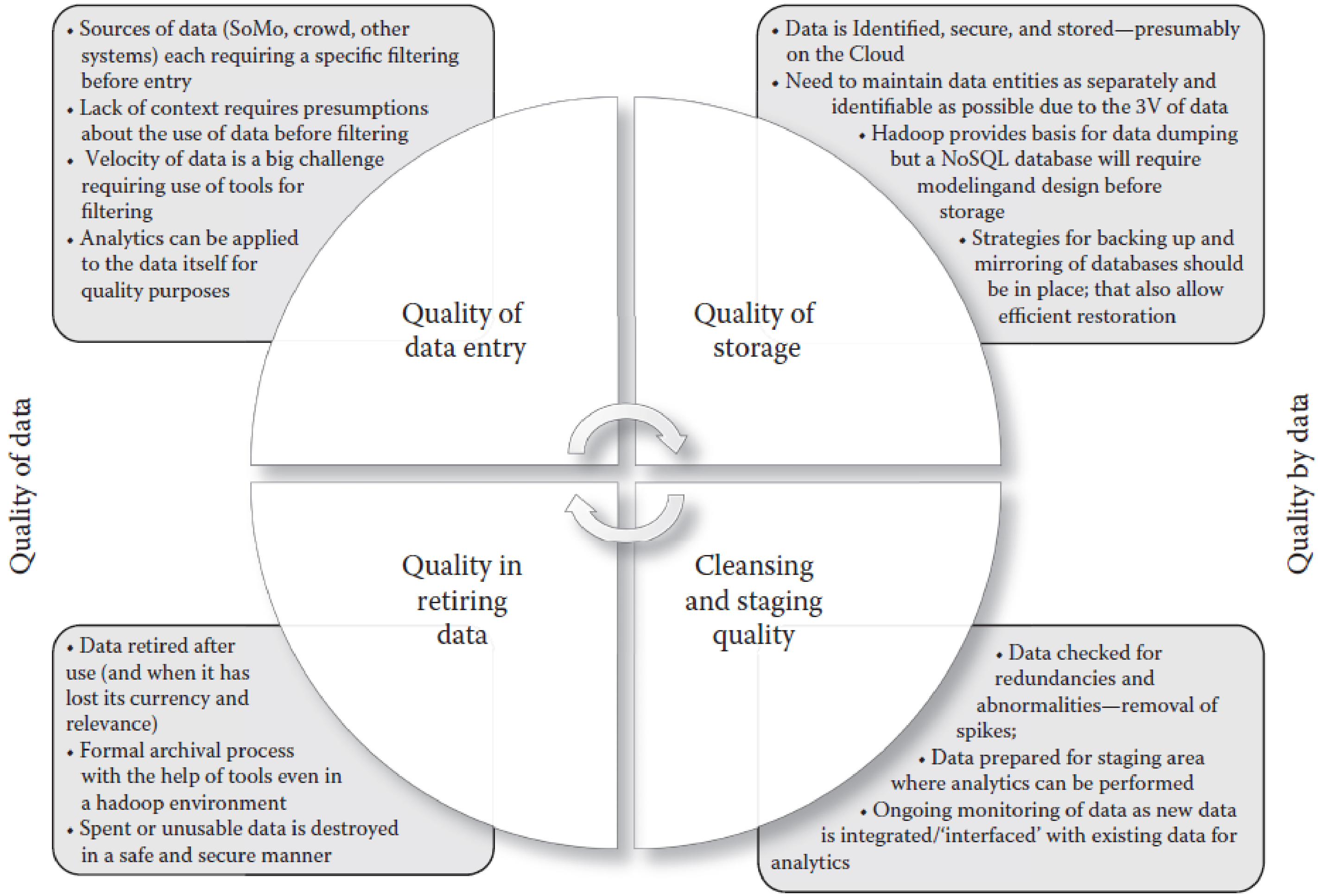
Design Science Research on Data Governance

Malgorzata Pankowska

Nice, July 30, 2019

[Unhelkar B. (2018) Big Data Strategies …]

Data Governance Quality Practices:

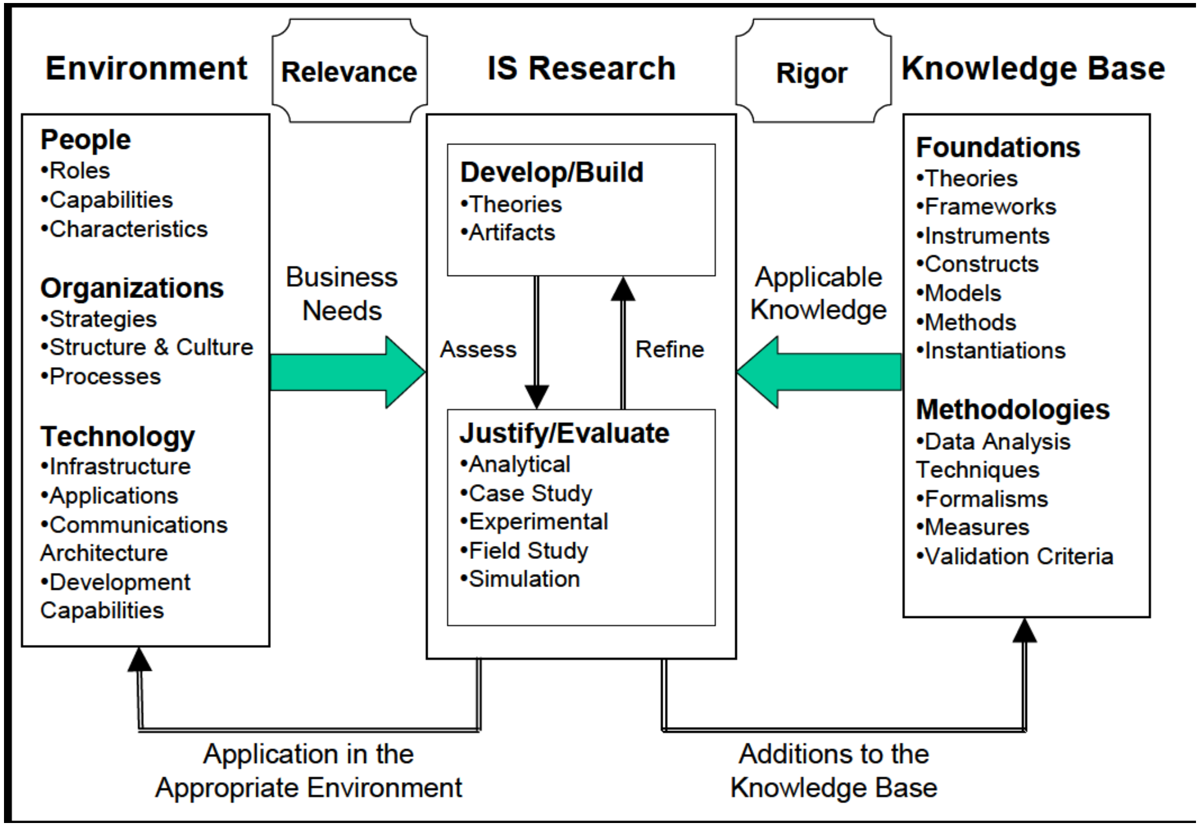


- Data profiling; cleaning and standardizing data; applying syntax, semantics, and aesthetic checks to data, algorithms and code quality
- Identification of source of data and tracing data,
- Using standard architectural reference models, controlling the business processes, continuous testing; using agile techniques to enable a high level of visibility and feedback; implementing governance policies for data

For Quality Preserving:

- Combining Design with Research
- Applicability of Design Science Research in Data Governance
- Focus on Quality in the whole Data Governance Process

- **Research for Design**
- **Research into Design**
- **Research through Design [Frayling(1993)]**

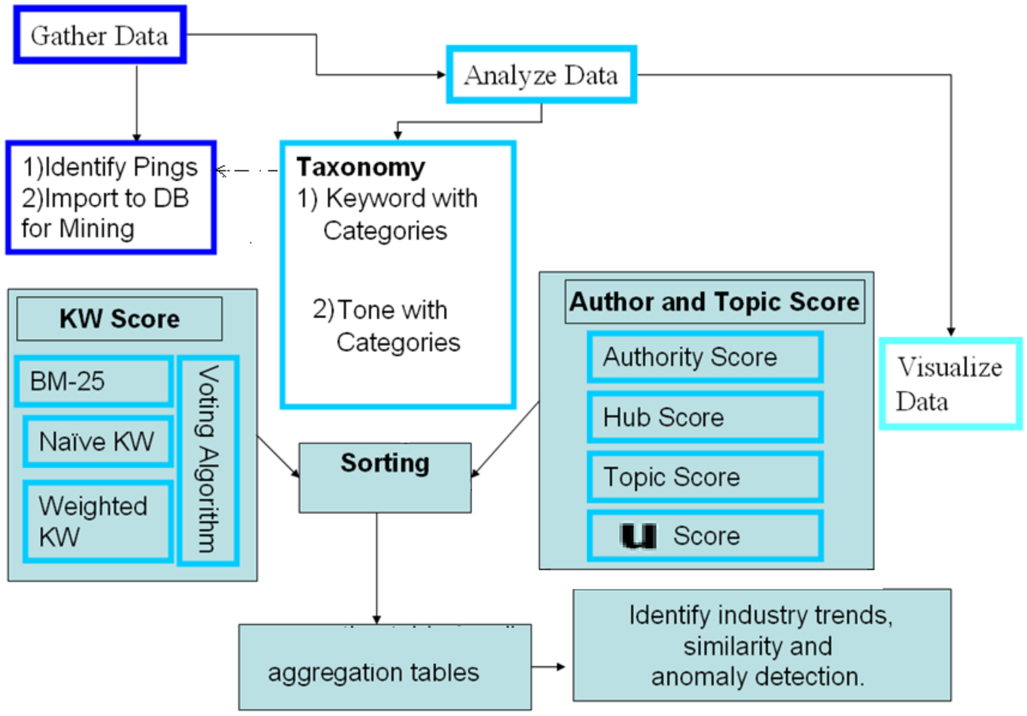


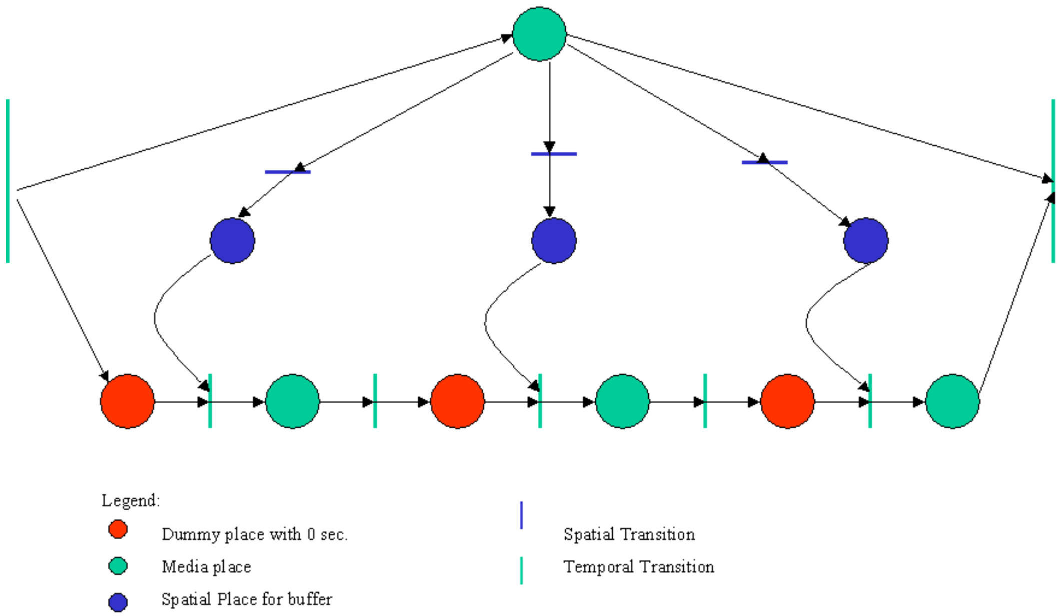
[Hevner A.R., March S.T., Park J (20004)]

Creating Summarized Actionable Insights From Tweets and Generating Meaningful Composite Multimedia Objects: lessons learned

Ahmed Gomaa, PhD
University of Scranton

Lessons learned from analyzing feed and Creating Meaningful Composite Multimedia Objects





Lessons Learned

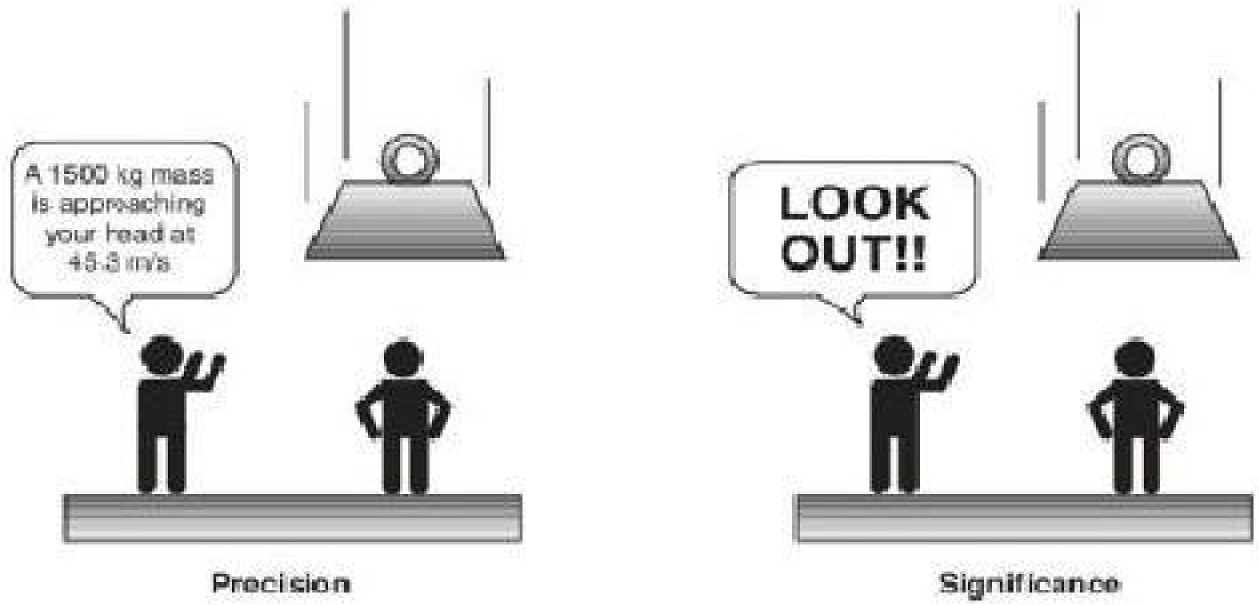
- Existing tools do not fully follow published standards.
- Existing tools do not fully provide the customization capabilities to enable the implementation of new theories.
- Expectations of cleaning the data at a 100% rate is very expensive. It is more reasonable to focus on how much better the results are.
- An 80/20 approach seems to be the most reasonable, where you get 80% of the results at a 20% of the cost.
- Improvement will always happen, but in phases, given that technology is evolving.

Precision vs Significance: which is the meaning of Information Quality?

Nice, July 2019

Mauro Dell'Orco – Polytechnic University of Bari (Italy)

Precision vs Significance: which is the meaning of Information Quality?



Source: MATLAB® Fuzzy Toolbox

In 1989, Smithson said:

"Western intellectual culture has been preoccupied with the pursuit of absolutely certain knowledge or, barring that, the nearest possible approximation of it."

This preoccupation is responsible for the dominance of precision versus significance in the Western science, although precision not always makes more sense than significance. In the left side of this figure, a guy is very precise, but the guy in the right part of the figure is more effective, simply shouting "look out!"

Precision vs Significance: which is the meaning of Information Quality?

Principle of Uncertainty Invariance:

$$H = -\sum_{i=1}^{n} P(x_i) \cdot \log_2 P(x_i)$$

$$U = \sum_{i=2}^{n} log_2(i) \cdot [Poss(x_i) - Poss(x_{i+1})]$$

H =U

$$P(x_i) = \frac{Poss(x_i)^{\gamma}}{\sum_{i=1}^{n} Poss(xi\ )^{\gamma}}$$

The idea is that Uncertainty, understood as the counterpart of Information, should be the same, regardless of the mathematical framework with which it is calculated. We have two Uncertainty measures, one probabilistic (i.e. the Shannon Entropy, H in the figure) and another one derived from Possibility Theory (Zadeh, 1978), called U-Uncertainty (U in the figure) and introduced by Klir and Folger in 1988. By solving numerically these equations, in which only γ is unknown, we obtain the value of γ and then we can calculate the probability distribution that preserves Uncertainty and thus the quality of Information.