Assessment Methods for Information Quality Criteria

Felix Naumann[*]
Humboldt-Universität zu Berlin
Unter den Linden 6
D-10099 Berlin
Germany
naumann@dbis.informatik.hu-berlin.de

Claudia Rolker
Forschungszentrum Informatik (FZI)
Haid-und-Neu-Str. 10-14
D-76131 Karlsruhe
Germany
rolker@fzi.de

Abstract

Information quality (IQ) is one of the most important aspects of information integration on the internet. Many projects realize and address this fact by gathering and classifying IQ criteria. Hardly ever do the projects address the immense difficulty of assessing scores for the criteria. This task must precede any usage of criteria for qualifying and integrating information.

After reviewing previous attempts to classify IQ criteria, in this paper we also classify criteria, but in a new, assessment-oriented way. We identify three IQ criterion classes, each with different general assessment possibilities, which we discuss in detail. For several exemplary criteria we give detailed assessment methods. Finally, we consider confidence measures for the methods. Confidence expresses the accuracy, lastingness, and credibility of the individual assessment methods.

1 Introduction

Low information quality is one of the most pressing problems for consumers of information that is distributed by autonomous sources. This is true both for casual users of WWW information services and for decision makers using an intranet to obtain data from different departments. The need for measures against low quality is clear and many projects have proposed methods to enhance information quality and data quality respectively. However, most approaches lack methods or even suggestions on how to assess the quality scores in the first place.

IQ assessment is rightly considered difficult for several reasons:

[*]This research was supported by the German Research Society, Berlin-Brandenburg Graduate School in Distributed Information Systems (DFG grant no. GRK 316).

1. IQ criteria are often of subjective nature and can therefore not be assessed automatically.

2. Information sources usually are autonomous and often do not publish useful (and possibly compromising) quality metadata. Additionally, many sources take measures to hinder IQ assessment.

3. The enormous amount of data to be assessed impedes assessment of the entire information set. Thus sampling techniques are often necessary which decrease the accuracy of the assessed scores.

4. Information from autonomous sources is subject to sometimes surprising changes in content and quality.

The most reliable source for IQ scores would be the information sources themselves. However, such IQ metadata is most often not made available, especially if the source is in competition with other sources. Therefore, methods must be developed, that independently assess IQ metadata in an efficient manner. I.e., assessment should be as automated as possible but as user guided as necessary.

This paper identifies three sources for IQ metadata, namely the user, the source, and the query process. The three metadata sources correspond to three classes of assessment methods. This paper reviews existing classifications and develops the new assessment-oriented classification. For each class we discuss general IQ assessment issues and give specialized examples for a comprehensive set of IQ criteria. Finally, we examine class-specific issues concerning confidence in the assessed IQ scores.

1.1 Related Work

Several research projects have tackled the problem of assessing scores for information quality criteria.

Wang et al. present an information quality assessment methodology called AIMQ which should help organisations to assess the status of their organisational information quality and monitor their IQ improvements over time [WSKL99]. AIMQ consists of three components. The first component is the Product-Service-Performance model which divides their set of IQ criteria into four classes. From this model a questionnaire—the second component—of 65 assessment items, some demographic questions, and space for comments is developed. The questionnaire should be sent to different organisations and should be answered by all respondents within an organisation. The respondents are asked to focus their answers on one specific set of information that is of importance to their organisation. The third component of AIMQ consists of two analysis techniques, one comparing the questionnaire results of different stakeholders of an information manufacturing system, and one comparing the questionnaire results of an organisation to that of a best practices organisation. Both techniques are executed on each IQ criterion class separately.

Bobrowski et al. present a methodology to measure data quality within organisations [BMY99]. First, a list of IQ criteria must be set up. These IQ criteria are divided into directly and indirectly assessed criteria. Scores for the indirectly assessed IQ criteria are computed from the directly assessed IQ criteria. In order to assess the direct criteria the traditional software metrics techniques are applied. These techniques measure data quality following the goal question-metric methodology. For each directly assessed criterion, a question is set up that characterizes the criterion and then a metric is derived to answer this question, giving a precise evaluation of the quality. From these metrics a user questionnaire is set up which is based on samples of the database.

Both AIMQ and the approach of Bobrowski et al. rely on questionnaires to find IQ scores. While this assessment method is inevitable for some criteria, it is by no means the only choice for all criteria. For instance, an automated method will be much more precise in assessing the average response time of a source. Why should the price of some information be determined by a questionnaire? Many techniques have been proposed to assess data accuracy. Our paper determines, which criteria can be assessed automatically and which criteria must determined by hand or by a questionnaire.

In [MR98] Motro and Rakov address two specific criteria—soundness and completeness of information sources. The authors propose automated assessment methods based on sampling of the source. Even though they are presented as an algorithm, they also rely on human input to verify whether some information is correct or not. Gruser et al. address in detail another criterion—response time [GRZZ00]. The authors suggest a prediction tool that learns response times of WWW information sources under several dimensions such as time of day and quantity of data.

1.2 Structure of the paper

In Section 2 we present existing classifications for IQ criteria and describe their philosophy. Furthermore, we present our assessment-oriented classification for IQ-criteria and classify a comprehensive set of IQ-criteria according to this classification. In Section 3 the assessment methods for each assessment class are examined in detail. Section 4 analyzes each assessment class with respect to its credibility, meaning, and general validity. The paper concludes in Section 5 and includes a definition list of IQ criteria in the Appendix.

2 Classification of IQ-Criteria

Many attempts have been made to compile and classify information quality criteria. In this section we review these attempts and identify three types of classification, all of which give no hints towards assessment methods for these criteria (Section 2.1). Thus, after the review we come up with a classification of our own, which classifies criteria according to the possible sources of the criteria scores (Section 2.2).

2.1 Existing Classifications

In [NR99] we compiled a list of information quality criteria taken from different projects that analyze information quality. Here we discuss these projects again, but pay special attention to their classification attempts and to the classification of [NR99]. We have identified three different kinds of classifications: semantic-oriented, processing-oriented, and goal-oriented classifications.

We call a classification semantic-oriented if it is solely based on the meaning of the criteria. This classification is the most intuitive when criteria are examined in a most general way, i.e., separated from any information framework. A classification is processing-oriented if it partitions IQ criteria according to their deployment in different phases of information processing. Finally, a classification is goal-oriented if it matches goals that are to be reached with the help of quality reasoning. The following reviews and classification of IQ projects are summarized in Table 1.

- TDQM: Total Data Quality Management is a project aimed at providing an empirical foundation for data quality. Wang and Strong have empirically identified fifteen IQ criteria regarded by data consumers as the most important [WS96]. The authors classified their criteria into the classes "intrinsic quality", "accessibility, "contextual quality", and "representational quality". The classification is based on the semantic of the criteria. It is of use to describe the criteria but not to assess them. We call this classification *semantic-oriented*.

- MBIS: The criteria of the mediator-based information system (MBIS) of [NLF99] are based on the TDQM criterion set. However, the criteria were re-classified to adapt to the query planning processing steps. For the source-selection phase source-specific criteria are employed. For the planning phase where views are combined, view-specific criteria are employed. Finally, when presenting the information, attribute specific criteria are used. We call this classification *processing-oriented*.

- Weikum: In [Wei99] the author developed a classification of IQ-criteria in a visionary way: He distinguishes system-centric, process-centric, and information-centric criteria. Even though the author had an application-specific classification in mind, we call the classification *processing-oriented* because the three classes can be related to query processing steps distinguished in the preceding item.

- DWQ: The Data Warehouse Quality (DWQ) project is based on the criteria of TDQM [JV97]. The authors define operational quality goals for data warehouses and classify the criteria by the goal they describe. These are accessibility, interpretability, usefulness, believability, and validation. We call this classification *goal-oriented*.

- SCOUG: In [Bas90] the framework for judging the quality and reliability of databases in terms of their design, content and accessibility is described. The authors give a list of quality criteria with no special classification. But all criteria are described from an *goal-oriented* point of view with an exact description of their assessment.

- Chen et al.: In [CZW98] the authors give a list of quality criteria with no special classification. However the paper is heavily biased towards time-oriented criteria such as response time and network delay. Thus, their approach is *goal-oriented.*

- Requirement survey: The survey in [NR99] compiles IQ criteria from all of the previous projects and finds a classification of its own. The classes are content-related, technical, intellectual, and instantiation-related criteria, and thus, they are *semantic-oriented.* This classification helps to understand the nature of the IQ criteria but we will show that an assessment-oriented classification is orthogonal.

Project	Classification
TDQM	Semantic-oriented
MBIS	Processing-oriented
Weikum	Processing-oriented
DWQ	Goal-oriented
SCOUG	Goal-oriented
Chen et al.	Goal-oriented
Requirement survey	Semantic-oriented

Table 1: IQ-criterion sets with classifications

The mentioned classifications were undertaken with different goals in mind. As argued before, most projects have avoided the difficult issue of quality assessment or have only touched it briefly. The goal of this paper is to find a new assessment-oriented classification. Such a general classification is necessary to discuss assessment issues in an ordered manner, and also to guide creators of assessment methods in establishing new methods for possibly new criteria. The following section identifies three classes that partition IQ criteria by the possibilities to assess their scores.

2.2 Three IQ Classes

The quality of information is influenced by three main factors: the perception of the user, the information itself, and the process of accessing the information. The three factors can be seen as the subject, object and predicate of a query. Each factor is a source for IQ metadata, i.e., for IQ criteria scores.

- The user: Arguably, the user is the most important source for IQ metadata. Ultimately, it is the user who decides whether some information is qualitatively good or not. Users can provide valuable input, especially for extremely subject criteria like understandability.

  Existing assessment methods *solely* rely on users to provide IQ scores. At the same time, obtaining user input is time consuming and at times even not possible. We will argue that this user input is only necessary for some criteria.

- The source: For many criteria the information source itself is the origin of IQ scores. Often the sources supply criterion scores, voluntarily such as the price or involuntarily such as the completeness. Since the source provides information, it automatically provides metadata that can be used for IQ scores.

- The query process: Finally, the process of accessing the information is a source for IQ scores. Criteria such as response time can be automatically assessed without input from the user or from the information source.

The three sources for metadata correspond to three assessment-oriented IQ criteria classes as shown in Figure 1. We distinguish the three classes below and give an example for each.

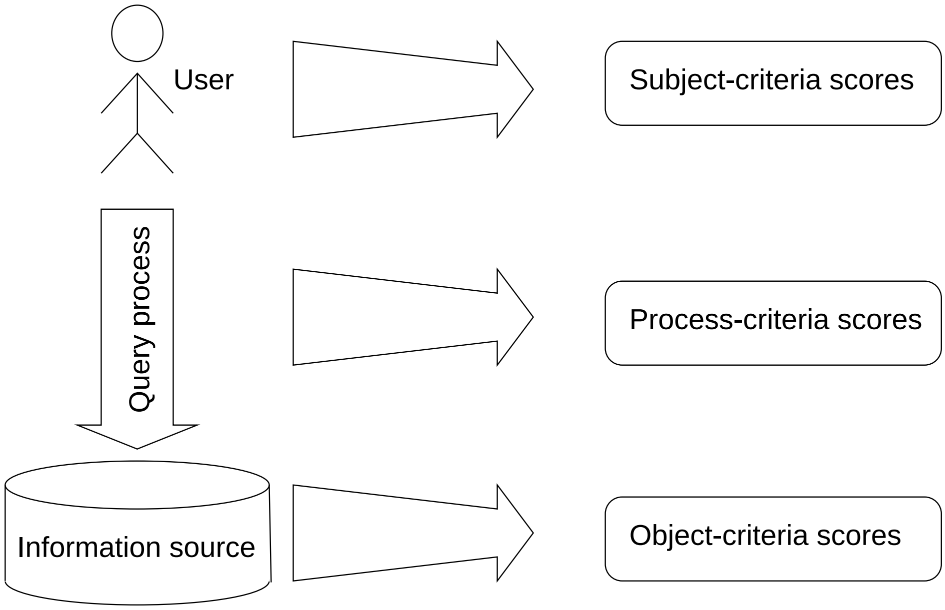

Figure 1: Three sources of IQ criterion scores

Table 2 lists a comprehensive set of IQ criteria within their class. These IQ criteria are taken from [NR99] where we unified the IQ criteria from several IQ criteria lists. A definition for each of these IQ criteria is given in the appendix. Table 2 not only classifies the IQ criteria according to our assessment classes but also provides the special assessment method for each criterion. We explain these methods in more detail in Section 3.

Subject-criteria. Information quality criteria are subject-criteria, if their scores can only be determined by individual users based on their personal views, experience, and background. Thus, the source of their scores is the individual user. Subject-criteria have no objective, globally accepted score. A representative subject-criterion is understandability.

Object-criteria. The scores of object information quality criteria can be determined exactly by a careful analysis of information. Thus, the source of their scores is the information itself, and hence, it is an object-criterion. A representative object-criterion is completeness.

Process-criteria. The scores of process-criteria can only be determined by the process of querying. The source of the scores are the actual query process. Thus, the scores cannot be fixed but may vary from query to query. The scores are objective but temporary. A representative process-criterion is response time.

Assessment Class	**IQ Criterion**	**Assessment Method**
Subject-Criteria	Believability	user experience
	Concise representation	user sampling
	Interpretability	user sampling
	Relevancy	continuous user assessment
	Reputation	user experience
	Understandability	user sampling
	Value-Added	continuous user assessment
Object-Criteria	Completeness	parsing, sampling
	Customer Support	parsing, contract
	Documentation	parsing
	Objectivity	expert input
	Price	contract
	Reliability	continuous assessment
	Security	parsing
	Timeliness	parsing
	Verifiability	expert input
Process-Criteria	Accuracy	sampling, cleansing techniques
	Amount of data	continuous assessment
	Availability	continuous assessment
	Consistent representation	parsing
	Latency	continuous assessment
	Response time	continuous assessment

Table 2: Classification of IQ Metadata Criteria

3 Assessment Methods for IQ-Criteria

In this section we first discuss difficulties of assessing scores for information quality criteria. Next we introduce some general assessment methods which we finally substantiate in some examples.

3.1 Accuracy vs. Practicality

Assessing IQ criteria is a difficult task. Assessment should be as accurate as possible but also as practical as possible. This is a conflict of goals and a compromise is difficult to achieve. Inaccurate assessment can either result in retrieval of low quality information or can lead to to avoidance of high quality information. Impractical assessment can either result in inaccurate assessment or can lead to undue assessment time and cost.

- Accuracy: An IQ score should reflect reality as accurately as possible. The first problem arises with the definition of the criterion. Only an accurately defined criterion can be assessed accurately. Further problems are distinct to the criterion class:
  - Subject-criteria: Scores for subject-criteria are only accurate for individual users, never for an entire group. Another obstacle is the amount of time a user will sacrifice for IQ assessment. The more time a user spends assessing different criteria, the more accurate the scores will be.
  - Object Criteria: The accuracy of object-criteria is particularly vulnerable to changes in layout and format of the information source. Also, due to the size of many sources, sampling techniques must be used. Their accuracy strongly depends on the sampling technique itself and the sample size.
  - Process: Scores of process-criteria are especially prone inaccuracy. Typically their accuracy declines over time and they are the most accurate at the time they were determined.
- Practicality: An assessment method should be as practical as possible. Inscrutable algorithms are neither trusted by a user nor easy to maintain. Any assessment method should by understood by the user and should be easy to adapt to new sources and new requirements.
  - Subject-criteria: As noted earlier, users will not spend much time on source quality assessment. A simple questionnaire must be enough, possibly with default scores. If a user changes her mind about the assessment of a source, an update of the scores must be as practical as possible as well.
  - Object-criteria: Assessing object-criteria should neither be too costly nor too time consuming, especially if the methods must be applied on a regular basis to keep the up-to-date.
  - Process-criteria: For process-criteria, the same arguments apply as for object-criteria, and even more so. Process-criteria are—by definition—assessed during a query process. If this is to time-consuming, the user will not be satisfied.

3.2 Score Units and Ranges

To correctly and usefully assess IQ criteria, system designers and users must agree on a *unit* to measure the criterion and on a *range* within which the scores may lie.

- Subject-criteria: It is often difficult to identify units for subject-criteria. For instance, understandability does not have an obvious unit, other than some grade. If there is some unit, great care must be taken when assigning and defining it: Since typically a user will assess subject-criteria, the units for these criteria must be intuitive, uncomplicated, and well described. Only then, will the user be able to assign proper and appropriate scores.

  Also, the range of the criterion scores must be clear to the user. Typical ranges are from 1 to 10 or a percentage. If these are not known to the user, the scores may be askew.

- Object-criteria: For some criteria expert input is needed. Thus, the same considerations as for subject-criteria apply. For other criteria such as price unit and range must only be agreed upon once and are clear from then on.

- Process-criteria: The unit and range for process criteria is usually non-ambiguous: it is derived from the criterion itself. Time-related criteria like response time or latency are measured in seconds, availability is a percentage, etc. From these units the ranges can also be derived in an unmistakable manner. Seconds range from zero to infinity, percentages are between 0 and 100, etc.

3.3 Assessing Subject-Criteria

As defined in Section 2.2, subject-criteria must be assessed by the individual user. In consequence, they are specific to each user, i.e., an information system must keep individual IQ score profiles for each user for all subject criteria. When assessing subject-criteria it is especially important to

- supply the user with an exact definition of the criterion she is assessing. The definition should be short, comprehensible, and non-ambiguous. The definition can be made up of several subcriteria; for instance, to define Understandability the subcriteria language, structure, and graphical layout can be mentioned to guide the user.

  If the definition is not given, a user may confuse two criteria, give imprecise scores, or assess wrong aspects of the source.

- give the range the score should be in. Section 3.2 discusses problems of communicating the range.

- provide examples of typical good and bad cases to guide the user. These quality of the examples should be especially visible in the certain criterion to be assessed.

The only way a system can support assessment is by providing default values as guidelines to users. First, if the user is not willing or able to provide individual score, the default scores can be used. Second, the default scores are somewhat a guideline for users. Either a system administrator provides the default scores, or the average score of other users is given.

In Table 2 we mention three methods of assessing subject-criteria—user experience, user sampling and continuous user assessment. All three methods should be supported by a well designed questionnaire.

- User experience: For the user experience method, the user must apply her experience and knowledge about the sources. This may include hear-say, experiences with the source itself, news reports, etc. For this method it is unnecessary to actually use the source or sample some information.

- User sampling: To apply this method the user must sample results of the information source[1]. Simply by looking at several results the user should be able to find an IQ score for the criterion to be assessed.

  This sampling must only be performed once in a while, either on a regular basis or when the source undergoes relevant changes. A system can support the process by suggesting a new assessment whenever appropriate.

- Continuous user assessment: Just as with user sampling, the user must sample the information by looking at it, reading it, or even by actually doing whatever she wanted to do with the information. However, continuous assessment analyzes every information received and not only samples.

  This method is by far the most time consuming and least rewarding of the three. However, it must be applied to criteria where the score of one information allows no prediction of future scores and where it is extremely difficult or even impossible to find representative samples of the information.

3.4 Assessing Object-Criteria

Object-criteria scores can be assessed mostly automatically—only an occasional user or expert input may be necessary. In a WWW information source setting, the scores of object-criteria can often be obtained by parsing the main page of the source. Also, scores for object-criteria are not often subject to change.

- Contract: For some criteria, the scores can be assessed by considering the terms of the contract (agreement) between the source and the information consumer. Usually, price and support are determined in some agreement. These terms can be valued by an expert who then assigns scores to the criteria.

- Parsing: Parsing a source is often a valuable tool to asses criteria. We distinguish structural parsing and content parsing. Structural parsing is discussed in the following section for process-criteria. Content parsing considers the actual information

[1]Finding appropriate and representative samples is a problem of its own and not covered in this paper.

and other content of the information and the information source. Aspects such as the presence of a documentation or customer support can be gained by searching the information for help links or the like. Aspects such as security of information can be assessed by analyzing the protocol by which the information is delivered.

- Sampling: Some object-criteria concern the entire content of the information source. To assess the accurate score, the entire content would have to be considered. To avoid this time-consuming and possibly costly task, sampling techniques can be applied. Sampling techniques choose a representative subset of the information and only consider those for quality assessment.

- Expert input: A human expert is needed to assess some criteria. The expert should follow some guideline to guarantee accuracy and comparability of the scores.

  Expert input is a method to assess object-criteria despite the fact that an expert is a human and thus prone to assess the scores subjectively. Object-criteria are named so due to the source of their scores, i.e., the object as explained in Section 2.2. But also, criteria that can be assessed only by expert input are still assessed objectively—merely a human expert is needed to find accurate scores.

- Continuous assessment: Some criteria can only be assessed by continuously checking how well the information source does in that criterion. This is true for the object-criterion reliability and also for many process criteria as we will discuss in the following section.

3.5 Assessing Process-Criteria

- Cleansing techniques: Accuracy or data quality has been subject of several research projects [HS98, MWS98, GFSS00]. The impact of data errors on data mining methods and data warehouses has given rise to data cleansing methods. The methods identify and eliminate a variety of data errors. The identification techniques can be used to count errors and thus to assess data quality.

- Continuous assessment: Several criteria underlie frequent changes. Some changes depend on time-related aspects. For instance, latency heavily depends on net load and this on the time of day. Other criteria like availability additionally depend on hardware and software aspects of the information source.

  Continuous assessment measures quality scores at regular intervals. Each new score is added to the history and statistical methods can provide accurate and timely quality scores. A simple statistical measure is the average score over the entire history. More sophisticated methods can additionally consider the aging of quality scores and add weightings to more recently assessed scores.

- Parsing: As explained in the previous section, we distinguish content-based and structural parsing. Structural parsing applies to process-criteria. It considers the structure of the information such as positioning of tables, presence of graphics etc.

4 Confidence in IQ Assessment Methods

When using the scores of certain IQ criteria, confidence in these scores must also be considered due to the way the scores were determined. It is important to notice that this awareness is independent of what the scores are used for (e.g., comparison of different information sources or consideration of only one information source) and what the goal of the use is (e.g., selection of an information source, finding out improvement potentials, or determining an overall—aggregated—quality score).

4.1 Basic Confidence

In order to gain a certain amount of confidence in the scores of an IQ-criterion, the indispensable presumption is that a detailed description of the assessment method and of the actual assessment implementation is available. This is true for any IQ-criteria, independent of the assessment method. Besides the full information about the assessment method, one also needs to know when the last assessment took place: Information sources tend to grow, change their appearance, revise their data gathering methods, etc. Most IQ scores age fast: the older they are, the less confidence they have.

These two kinds of information—assessment method and assessment date—are essential. If they are not available, confidence in the scores for the IQ-criteria has no basis and is assumed to be zero. Even if they are provided, full confidence in the values can never be gained because each class of assessment methods has its own uncertainties. In the following, we discuss for each class of assessment the sources of low confidence and its consequences.

4.2 Confidence in Subject-Criteria

The source of low confidence in subject-criteria scores can only be the person assessing the quality because the scores of subject criteria strongly depend on him/her. There are two kinds of uncertainties which might result in scores not representing the reality.

- Type-1 assessors: The person unintentionally assesses the quality too good or too bad due to the absence of knowledge and of experience. E.g., a manager does not work very often with an information source to be assessed, and inadvertently assesses the understandability of the information source very low.

- Type-2 assessors: The person intentionally assesses the quality too good or too bad due to personal or institutional aims. E.g., a department head assesses the understandability of an information source very high because he/she wants to show the top management the good quality of his/her department's work.

Usually, not only one person enters the scores for the subject criterion values, but many. This can be a homogenous or a heterogeneous group. In both cases Type-1 assessors and Type-2 assessors can take part in the assessment. So the knowledge about the homogeneity or inhomogeneity of the group does not influence the confidence in subject values. If only few persons determine the scores for the IQ criteria, then there is probably low confidence in these scores. If the group of assessors is large, then there is higher confidence because we assume that the scores of Type-1 assessors and of Type-2 assessors perish in the average value.

Low confidence in subject-criteria score can be fought by increasing user input, i.e., by information consumers having more influence in the scores and spending more time assessing them.

4.3 Confidence in Object-Criteria

In general, confidence in object-criteria is high due mostly due to simple verifiability. The only detriment could be too infrequent updates of the scores. We discuss confidence in object criteria in dependence of the used assessment method.

IQ scores determined by contract (e.g., price) gain high confidence as both parties (consumer and sales clerk) must respect them.

In the case of content parsing, there is high confidence if only the presence of a facility (e.g., online documentation) is evaluated. If the content must be parsed as well (e.g., completeness), the same sources for low confidence exist as for subject-criteria. Additionally, the parsing techniques are vulnerable to changes in the information appearance, making frequent assessment even more important.

Low confidence in scores determined by sampling exists if a non-representative part of the entire document source was taken for the computation. A proportion is non-representative if it is too small or the sample has been taken too long ago and the information source has changed since. If the proportion is representative, then there is high confidence in these scores.

If an expert is needed to determine the scores for certain IQ criteria, then similar sources for low confidence exist as for subject-criteria. However, we assume experts to be able to assess certain scores in a quite objective manner. For instance, an expert can assess verifiability simply by verifying some sample information. This can be done in an objective manner.

In case of continuous assessment, there is low confidence in its scores if the point of time when the assessment is taken is not representative. I.e., if the intervals are too large or the point of time is always at noon but not randomly determined. If the sets of point of times are representative, high confidence exists in these scores.

Low confidence in object-criteria scores can be fought by increasing assessment frequency and possibly by training expert assessors when necessary.

4.4 Confidence in Process-Criteria

The confidence in process-criteria must also be considered in dependence of the assessment method in use.

The source of low confidence in the cleansing techniques is the adjustment of the technique to find errors. A technique can be too sensitive and detect errors which do not exist. Or a technique is not sensitive at all and does not find the existing errors. Also, many cleansing techniques require user input—another source for diminished confidence. Many research projects examine cleansing techniques and all have some kind of success measure. Sophisticated techniques find up to 98% of all errors, thus confidence in accuracy scores is high.

For continuous assessment and for parsing confidence issues were discussed in Section 4.3. Low confidence in process-criteria score can be fought by repeating the query process often. Since the scores are assessed during the query process, confidence will rise, the more queries are issued.

4.5 Overall Sources of Low Confidence

In summary, the confidence in the scores of any IQ criteria is reduced either because of content-related uncertainties (subject criteria, expert assessment) or because of the uncertainty to believe in representatives (parsing, continuous assessment, sampling). Only the contract IQ-criteria and the cleansing technique do not fit into this bi-classification of overall sources of low confidence.

As a consequence, users will never have full confidence in a score of an IQ-criterion, but there are methods to raise confidence to an acceptable level.

5 Conclusion

In this paper, we presented our assessment-oriented classification for IQ-criteria. The classification is not based on the exact method the assessment is done, but on the entity/process which is the source of the assessed scores. We identified three IQ-criterion classes:

- assessment with respect to the user of information (subject IQ criteria)
- assessment with respect to the information source itself (object IQ criteria)
- assessment with respect to the query process (process IQ criteria)

We examined each class from the viewpoint of persons setting up the assessment and from the viewpoint of persons using the scores. Having the goal to come up with realistic scores, it turned out that each assessment class has its own problems and uncertainties. So, independent of the assessment class and the specific IQ criteria it is nearly impossible to have totally realistic and true IQ scores.

Nevertheless, it is very important to have a detailed description of the assessment available for the assessors and the users and to repeat the assessment regularly. It is desirable for the future to have a standard describing how and when the assessment should take place. Furthermore, the assessment (persons and tools) should be evaluated in order to ensure that the rules of the standard are fulfilled. A certified assessment similar to ISO 9000, which is used to have a world-wide unified quality management in design, development, production installation, and servicing in all application areas, simplifies the set up of assessments and increases the confidence and comparability in IQ scores.

References

[Bas90] Reva Basch. Measuring the quality of the data: Report on the fourth annual SCOUG retreat. *Database Searcher*, 6(8):18–24, October 1990.

[BMY99] Mónica Bobrowski, Martina Marré, and Daniel Yankelevich. A homogeneous framework to measure data quality. In *Proceedings of the International Conference on Information Quality (IQ)*, pages 115–124, Cambridge, MA, 1999.

[CZW98] Ying Chen, Qiang Zhu, and Nengbin Wang. Query processing with quality control in the World Wide Web. *World Wide Web*, 1 (4):241–255, 1998.

[GFSS00] Helena Galhardas, Daniela Florescu, Dennis Shasha, and Eric Simon. An extensible framework for data cleaning. In *Proceedings of the International Conference on Data Engineering (ICDE)*, San Diego, CA, 2000.

[GRZZ00] Jean-Robert Gruser, Louiqa Raschid, Vladimir Zadorozhny, and Tao Zhan. Learning response time for websources using query feedback and application in query optimization. *VLDB Journal*, 9:18–37, 2000.

[HS98] Mauricio A. Hernández and Salvatore J. Stolfo. Real-world data is dirty: Data cleansing and the merge/purge problem. *Data Mining and Knowledge Discovery*, 2(1):9–37, 1998.

[JV97] M. Jarke and Y. Vassiliou. Data warehouse quality design: A review of the DWQ project. In *Proceedings of the International Conference on Information Quality (IQ)*, Cambridge, MA, 1997.

[MR98] Amihai Motro and Igor Rakov. Estimating the quality of databases. In *Proceedings of the 3rd International Conference on Flexible Query Answering Systems (FQAS)*, Roskilde, Denmark, May 1998. Springer Verlag.

[MWS98] Steve Mohan, Mary Jane Willshire, and Charles Schroeder. DataBryte: A proposed data warehouse cleansing framework. In *Proceedings of the International Conference on Information Quality (IQ)*, Cambridge, MA, 1998.

[NLF99] Felix Naumann, Ulf Leser, and Johann Christoph Freytag. Quality-driven integration of heterogenous information systems. In *Proceedings of the International Conference on Very Large Databases (VLDB)*, Edinburgh, 1999.

[NR99] Felix Naumann and Claudia Rolker. Do metadata models meet IQ requirements? In *Proceedings of the International Conference on Information Quality (IQ)*, Cambridge, MA, 1999.

[Red96] Thomas C. Redman. *Data Quality for the Information Age.* Artech House, Boston, London, 1996.

[Wei99] Gerhard Weikum. Towards guaranteed quality and dependability of information systems. In *Proceedings of the Conference Datenbanksysteme in Büro, Technik und Wissenschaft*, Freiburg, Germany, 1999.

[WS96] Richard Y. Wang and Diane M. Strong. Beyond accuracy: What data quality means to data consumers. *Journal on Management of Information Systems*, 12, 4:5–34, 1996.

[WSKL99] Richard Y. Wang, Diane M. Strong, Beverly K. Kahn, and Yang W. Lee. An information quality assessment methodology. In *Proceedings of the International Conference on Information Quality (IQ)*, pages 258–265, Cambridge, MA, 1999.

A IQ Criteria

In this appendix we define all IQ criteria from Table 2 as we understand them. Of course, these definitions will not satisfy every situation or application and should be viewed as general proposals to avoid misunderstandings. Also notice that many criteria are similar to each other and typically not all criteria should be used at the same time. Rather, an application specific selection of criteria will help identify qualitatively good information and simultaneously reduce assessment cost.

After an brief description of the criterion we give a short list of synonyms that were used by various authors to express the same criterion. The synonyms were compiled from [Bas90, CZW98, JV97, NLF99, Red96, Wei99, WS96].

Availability Percentage of time an information source is "up". Also: accessibility, reliability, retrievability, performability

Accuracy Quotient of the number of correct values in the source and the overall number of values in the source. Also: data quality (as opposed to information quality), error rate, correctness, integrity, precision

Amount of data Size of result. Also: essentialness

Believability Degree to which the information is accepted as correct. Also: error rate, credibility, trustworthiness

Completeness Quotient of the number of response items and the number of real world items. Also: coverage, scope, granularity, comprehensiveness, density, extent

Concise representation Degree to which the structure of the information matches the information itself. Also: attribute granularity, occurrence identifiability, structural consistency, appropriateness, format precision

Consistent representation Degree to which the structure of the information conforms to that of other sources. Also: integrity, homogeneity, semantic consistency, value consistency, portability, compatibility

Customer support Amount and usefulness of online support through text, email, phone etc.

Documentation Amount and usefulness of documents with meta information. Also: traceability

Interpretability Degree to which the information conforms to technical ability of the consumer. Also: clarity of definition, simplicity

Latency Amount of time until first information reaches user. Also: response time

Objectivity Degree to which information is unbiased and impartial.

Price Monetary charge per query. Also: query value-to-cost ratio, cost-effectivity

Relevancy Degree to which information satisfies the users need. Also: domain precision, minimum redundancy, applicability, helpfulness

Reliability Degree to which the user can trust the information. Note: technical reliability is synonymous to availability.

Reputation Degree to which the information or its source is in high standing. Also: credibility

Response time Amount of time until complete response reaches the user. Also: performance, turnaround time

Security Degree to which information is passed privately from user to information source and back. Also: privacy, access security

Timeliness Age of information. Also: up-to-date, freshness, currentness

Understandability Degree to which the information can be comprehended by the user Also: ease of understanding

Value-Added Amount of benefit the use of the information provides.

Verifiability Degree and ease with which the information can be checked for correctness. Also: naturalness, traceability, provability