A FLEXIBLE QUALITY FRAMEWORK FOR USE WITHIN INFORMATION RETRIEVAL

(Research-In-Progress)

M S E Burgess, W A Gray, and N J Fiddian
Department of Computer Science, Cardiff University, Wales, UK.
{m.burgess, w.a.gray, n.j.fiddian}@cs.cardiff.ac.uk

ABSTRACT Within recent years the volume of data readily available to the information consumer has dramatically increased in size. Although plentiful, this information is also of varying levels of quality, with providers ranging from multi-national corporations to specialist societies, and professional individuals to students with limited knowledge. As such, it is becoming increasingly difficult when searching for information to find precisely what is required. The two hurdles that prevent the finding of relevant information are therefore ‘information overload’ and ‘information quality’.

Our proposed solution to this problem consists of the development of a methodology for using quality metrics as an aid to information searching. By providing the consumer with a facility for stating their information requirements in terms of quality, via a set of quality metrics with associated importance weightings, the precision of search results is significantly improved. This method also allows easy manipulation of the search criteria, as both the metric selection and importance weightings can be quickly and easily changed, giving each user the opportunity to experiment with the various quality metrics, and observe how different quality weightings affect the returned results.

1 INTRODUCTION

Within recent times the amount of information available to the consumer has exploded, with increased numbers of people having access to distributed information resources such as local Intranets and the Internet [9]. It is now possible for these consumers to search for information on a wide range of topics, such as news articles, research documents, statistical data, and information on products and services. This information may also be stored in a variety of formats, such as web pages and databases, supplied by various information producers.

As increased amounts of information becomes available, users begin to suffer from information overload [20] [21]. Hence it becomes increasingly difficult, or even impossible, to find the information they require. This information may well be included in the result set returned from a database query or Internet search. However, these results also often include so much irrelevant information, it becomes impossible to manually search through them all to identify the relevant information.

Information overload can also be a problem when a user is searching for information within a given dataset, without knowing precisely what they need. In such situations the user may create a general query to investigate the type of information available, only to find themselves inundated with information that is of little interest.

Once information on the required subject is found there is also then the problem of information reliability. This is particularly an issue when obtaining information from the Internet, where there is no restriction on the quality of information that can be published. Such distributed information environments promote the idea that anyone can share their information with everyone else. Information suppliers can therefore be as varied as students to multi-national corporations, meaning there is no control over the quality of the information that is placed into the public domain. This can be a serious problem when, for example, looking for information on prescription drugs [2]. It can also be a problem, although less serious, when researching topics for use in student-level research [15].

Our research is focused on challenging this problem by enabling the user to state their desired level of quality in terms of a query filter. This will automatically filter the result set returned by the user's query, thus tailoring the information to their individual requirements. Research is also currently being conducted into the combination of data from multiple sources, each with an overall rating of quality that can be used as a method of weighting information from the various sources.

Previous projects that have used quality attributes to filter information have only selected those attributes relevant to their particular area (for example, research journals [12]), hence each time a new system is developed a new set of quality metrics must be collated from scratch. Our research overcomes this restriction through the development of a generic hierarchical framework of quality that can be reused within any information retrieval (IR) system. Once created, this generic model can be augmented with domain specific frameworks, to provide a method of describing the desired level of quality within a chosen domain of interest.

In this paper, we present our flexible framework of quality, which builds on our previous work on creating a taxonomy of quality [6]. This is achieved by selecting the aspects of quality that are of most interest to the user, providing a rating for relative importance of these aspects, and then using this information to automatically generate a set of results that meets their requirements.

In section 2 of this paper we present a brief overview of research that has previously been carried out in the area of information quality, including previous attempts at generating a model of quality. In section 3 we then go on to describe our work into creating a domain-independent model of quality, which can be used to describe quality from the perspective of the information consumer. This is followed in section 4 with a discussion about how information quality ratings can be obtained from various sources, such as the information suppliers and user communities. Section 5 discusses our experiments in using quality metrics within an information search system. Section 6 then discusses future work, and the paper concludes in section 7.

2 RELATED WORK

The idea that quality is a multidimensional entity is not new. Multiple quality dimensions have been identified in projects covering a diverse range of subject domains, including software engineering [3;8], engineering products [23], enterprise modelling [19], query processing [22] and total data quality management [18;24;29]. The research conducted in these, and other, projects has resulted in definitions of quality ranging from collections of potential quality metrics to hierarchical models. The main commonality among them however is the investigation of quality from an organisational perspective, or looking into how the quality of existing data can be improved. The research discussed in this paper differs from this standpoint in that we are investigating quality of information from a user perspective:

how can the user get information of the best available quality, from data sources over which they have no control?

The most widely accepted of the currently available models of quality is a hierarchical framework created by Wang and Strong [30], containing 15 dimensions of quality organized into 4 categories: intrinsic data quality (DQ), contextual DQ, representational DQ, and accessibility DQ, as in Table 1. Although this model of quality contains 118 quality attributes, their work mainly focuses on the use of quality dimensions (groups of attributes) to model quality of existing data.

DQ Category	**DQ Dimensions**
Intrinsic DQ	Accuracy, objectivity, believability, reputation
Contextual DQ	Relevancy, value-added, timeliness, completeness, amount of information
Representational DQ	Interpretability, ease of understanding, concise representation, consistent representation
Accessibility DQ	Access, Security

Table 1: Wang and Strong's DQ Categories and Dimensions [30]

One of the major differences between our model of quality and that proposed by Wang and Strong is the level of detail. Unlike their model, we concentrate on both the dimension and attribute levels, leaving the user to make the final decision regarding the level of detail they desire.

For example, in Wang and Strong's model there exists a dimension called 'value-added', which within our model is a high level category of quality, with three subcategories and twenty-two quality attributes. Our model therefore augments theirs by building on their quality dimensions, and increasing the choice of qualities available to the user. Our research focuses on evaluating the usefulness of the fine-tuning of an information search that this level of detailed quality requirements provides.

3 MODELLING QUALITY

Quality is not an easily definable term, as it is not absolute. It has many different aspects and its meaning varies across different situations and users [4] [11]. A formal method is therefore needed to describe this term, freeing users from the need to produce detailed quality requirements when requesting information, but still allowing them to stress the importance of various desirable characteristics, into a definition of their required level of quality.

Although previous attempts have been made to create a model of quality, as illustrated earlier, few of these have focussed on the information consumer perspective (the principal exception to this being Wang and Strong's work, where the aspects of quality were derived from data consumer questionnaires). The first aim of this project was therefore to create a model of quality that could be used by an information consumer, to describe the level of quality they desire from the information they seek, without needing to concern themselves with the data that is actually available.

3.1 Hierarchical Framework

During the earlier part of this research project we developed a hierarchical taxonomy of quality, containing a generic collection of quality categories and attributes. This taxonomy was created by combining both a user perception of the term quality and domain specific interpretations, incorporating aspects of quality discussed in a variety of domains, including those discussed in section 2 of this paper.

As the meaning of quality varies across different requirements, users and domains a static model of quality is inappropriate, as it needs to be adaptable according to the current situation and perspective of the user. This is particularly important in dynamic information domains, such as the Internet, where the available information is constantly changing. The quality attributes within our taxonomy were therefore categorised to enable a user to follow an intuitive path through the taxonomy, to find and select the attributes they need in their current situation. By allowing users to state the importance of the available quality metrics, either explicitly or implicitly by monitoring the way the taxonomy is used, and allowing the taxonomy to be updated, its structure and contents will change over time. This will increase the usefulness of the taxonomy as it adapts to reflect the current consumer needs.

An important difference between our model of quality and previous attempts, such as that produced by Wang and Strong [30], relates to its flexibility. Within previous models the quality metrics are static – dimensions remain at the dimension level, with each consisting of a set of unmoveable metrics. As our model is based on unpredictable user requirements the need for flexibility is increased. Our framework therefore allows for alterations within the hierarchical structure, with metrics being able to move, both between categories and within the hierarchical structure (e.g., a bottom level quality metric can be changed to a dimension). This allows the development of our framework over time as it is used, to maintain its relevance to the information consumer.

It will also be possible to update this quality framework as it is used over time, by monitoring those quality metrics that are seen as most important, and modifying the framework accordingly. This revision process can either be done automatically through user preference monitoring, or by enabling the user to add to the framework, and remove those attributes that are no longer needed. This will be particularly useful if applied within a user community, where this revision can be based on all members' usage patterns, and shared amongst the group. To demonstrate the ease with which framework edits can be conducted manually, a Quality Toolkit was developed as part of this project, which is discussed further in section 5.1.

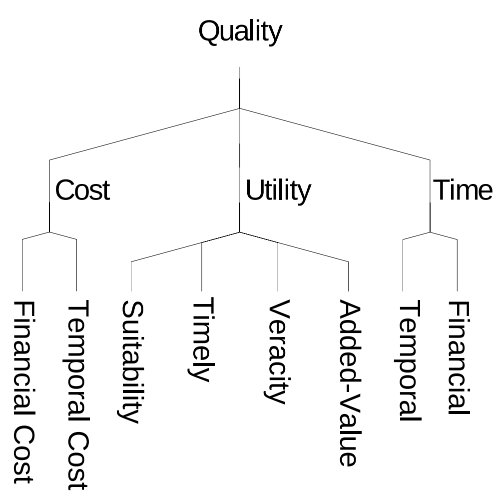


Figure 1: Upper Level of the Generic Framework of Quality

An outline of our quality framework can be seen in Figure 1, which shows the upper level quality dimensions of the hierarchical taxonomy. Each of the leaf nodes in this figure contains either a collection of quality attributes, or further sub-dimensions of quality. For more detailed information about this taxonomy, and an in depth discussion of its construction, refer to Burgess, Gray and Fiddian [6]. The most recent version of this taxonomy can also be found on our quality research project web site [25].

This taxonomy is the basis for our generic quality framework: a formal description of quality that can be used within an application to facilitate the extraction of a personalised quality definition from an information consumer, and that can be developed further into domain-specific frameworks.

3.2 Quality Framework Representation

To enable our model of quality to be used within an IR system it is currently organised within a relational database, allowing the storage of all currently identified quality metrics, plus supporting information. The entity-relationship diagram for the main tables within this database can be seen in Figure 2.

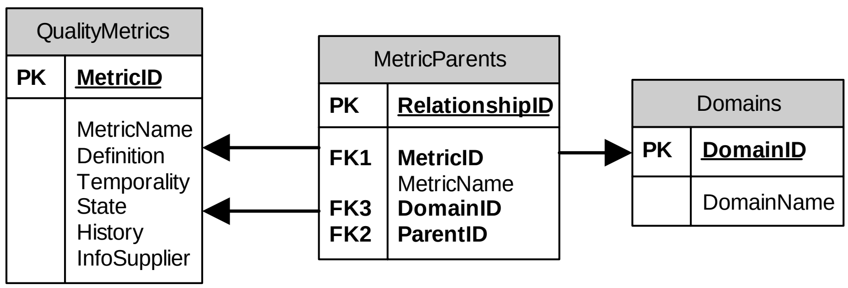


Figure 2: Entity-relationship diagram for the framework of quality

The information within these tables is as follows:

Quality Metrics

Definition	Contains details of each of the currently identified quality metrics.
MetricID	Unique reference number for each quality metric.
MetricName	Name of the quality metric.
Definition	A definition of the metric, used to provide assistance to the user when selecting the metrics that best suit their needs.
Temporality	Relates to the nature of each metric over time, stating whether they are static or dynamic over a period of time, for example 6-months.
State	A Boolean value stating whether the metric is positive or negative in nature. For example, the concept of *reputation* is positive as a high value is desirable, yet *price* is negative as a low value is perceived to be best.
History	States whether historical data is needed on a metric to be able to provide confident information regarding its value.
InfoSupplier	States from where information about that quality metric can be obtained, for example, the information producer, user feedback, or from independent third parties.

Domains

Definition	Used to store information about currently available subject domains, the models for which are discussed later in the paper, in section 3.4.
DomainID	Unique reference number for each subject domain.
DomainName	Name of the subject domain.
Extra information	This table also contains information on how to access the data sources used within this domain, such as addresses of databases within which the domain-specific information is stored.

Metric Parents

Definition	To record the hierarchical structure of each quality framework, by stating which metrics participate in parent-child relationships.
RelationshipID	Unique reference for each parent-child metric relationship.
MetricID	Reference number for the metric concerned.
MetricName	The name of that metric, recorded again here for clarity.
DomainID	Reference number for the domain in which this metric is relevant.
ParentID	Reference to another quality metric, which acts as the parent for this metric within the current hierarchy.

The last table, 'Metric Parents', was created to map the parent-child relationships within the framework, to allow for metric flexibility. For example, by storing this relationship in this manor moving metrics to different positions within the framework becomes a quick and simple process.

This table also eliminates the distinction between metrics and their parents (i.e., 'dimensions') so ensuring the flexibility of created quality frameworks, as all metrics can be moved within each framework, and can each be a child-level quality metric or a quality dimension.

The database method of representation was chosen for ease of use within our experimental applications, which have been developed to demonstrate our ideas. However, this need not be the final form of framework representation. Work is also currently underway into an XML model for the representation of quality frameworks. This will provide a more flexible representation method, due to its platform independence and ease of access when used within a distributed environment.

3.3 Quantifying Qualitative Quality Metrics

A number of the attributes within our quality framework are quantitative in nature so are straightforward to store and manipulate. For example, the attribute of 'currency' may be recorded as the date or time the information was last updated. Those that are qualitative, such as 'reputation' and 'accuracy' can also be made quantifiable by representing them as ordered sets of values, from which the required rating can be chosen. In the case of 'reputation' a value may be chosen from a scale of 1 to 10, where 1 indicates a very poor reputation, and 10 indicates excellent. By using this method of representation for qualitative quality metrics it becomes possible to define parameters by which each characteristic can be measured.

This quantification is possible as we are not concerned with absolute values of quality, but rather relative values.

3.4 Domain-Specific Quality Frameworks

Although we have presented a generic model of quality, it is not of great practical use on its own. This is because of its lack of subject-specific quality metrics, which are needed for applying this model within a real-world environment. Our model has therefore been extended by the inclusion of a set of domain-specific quality frameworks, each based on the generic framework but containing a set of quality metrics that are applicable to a particular domain. For example, if searching for information on cars domain-specific metrics would include, among others, fuel economy, maximum speed, and passenger capacity. Within the experimental application developed to illustrate our ideas (as discussed in section 5) the principal experimental domain was chosen to be UK Universities, which contains such quality metrics as research and teaching quality, reputation, and student completion rates, obtained from two independent sources. Although only one domain has currently been implemented, experiments within other subject domains are scheduled to be conducted within the near future, to broaden our experience of domain implications and widen the testing of our work.

The university quality framework, as used within our demonstration system, can be seen in Figure 3. It shows the current state of this framework, but due to the flexibility of these models it is liable to change many times within its lifetime. These changes may include new metrics being added, unused ones being removed, and current metrics being moved between dimensions.

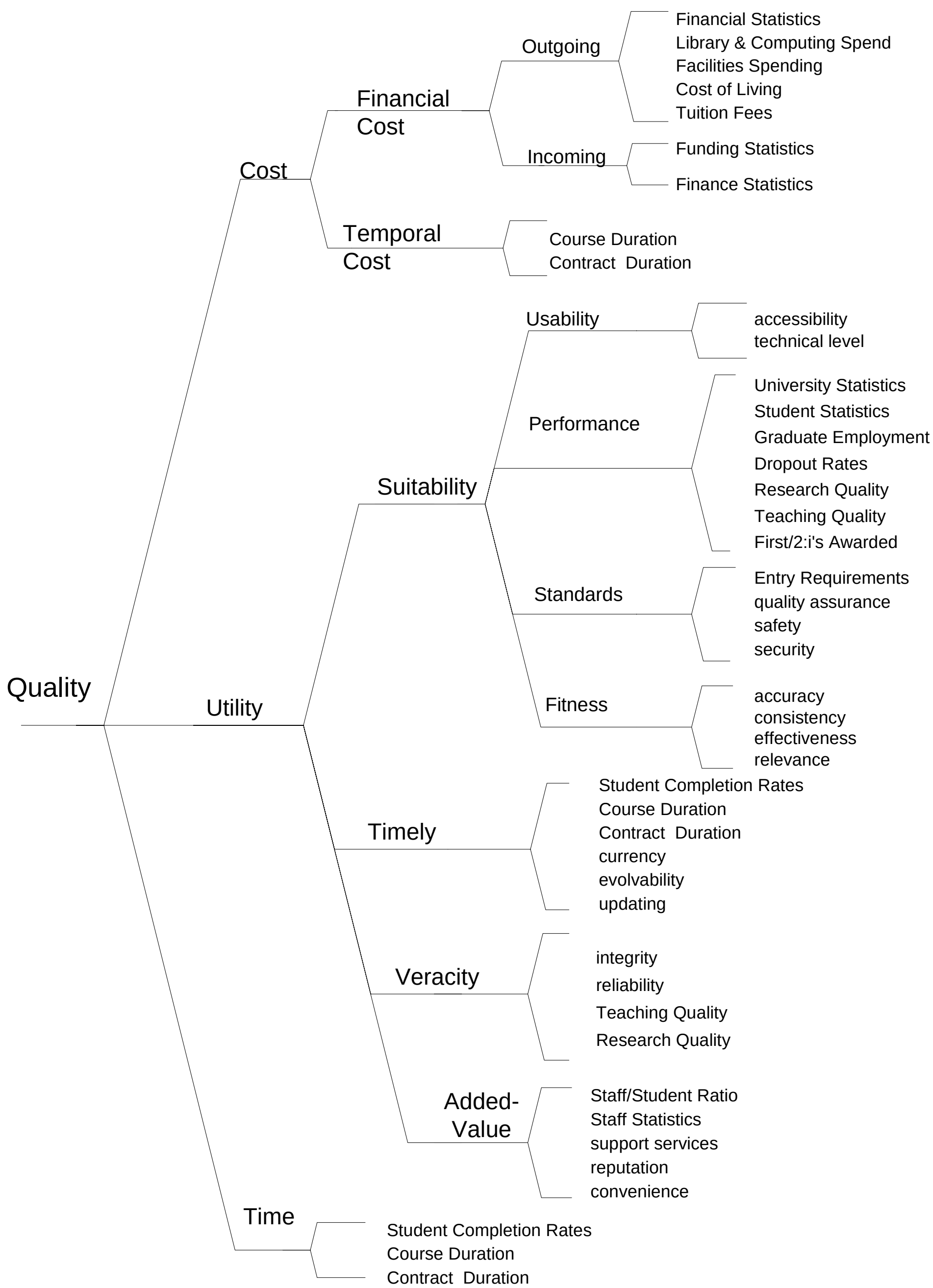


Figure 3: UK University domain framework of quality

As can be seen in this diagram some metrics are repeated within the framework. For example, *Research Quality* appears in both the dimensions of *Veracity* and *Performance*. This illustrates that dimensions of quality are not necessarily discrete. Metrics should be placed where they are relevant, and as they can be relevant in a number of places they can appear many times. This is because the aim of these frameworks is to make it easy for the user to describe their quality preferences, so a user should be able to find the required metrics by following an intuitive path through the framework, rather than having to look at all possible paths to find the metrics they desire.

Although allowing metrics to be placed more than once within each framework gives the impression of duplication, in practice this is not the case. Metric duplication only occurs in the *Metric Parents* table, illustrated in Figure 2. The actual metric is not duplicated. This means that if any alterations are made to one instance of, for example, the *Research Quality* metric, such as mapping information, this will also apply to all other instances of that metric within the framework.

4 ASSESSING QUALITY OF INFORMATION

Before being able to rate information based on it's level of quality a method for obtaining this knowledge must be developed.

There are two main levels at which quality of information can be assessed:

- Quality of the information source – Coarse Grained
- Quality of the data within a data set – Fine Grained

The ideal scenario is where both levels of quality are available for each data source. The fine-grained quality measure can then be weighted according to the coarse-grained measure, for each data set. This would be particularly useful when searching for information within multiple sources of varying levels of quality.

There are a number of potential methods for obtaining information about data quality. The information suppliers may be willing to provide quality ratings on the information they provide, based on such attributes as cost, availability, creation date and information currency. This type of information is likely to be more reliable than information supplied by them regarding, for example, their reputation. To obtain objective measures for these types of quality attributes, feedback is needed from either multiple users or from independent third parties, such as Which [31]: a UK-based consumer watchdog responsible for independently testing products and services.

Even if the supplier does not explicitly provide information regarding the quality of their information or product, it may be possible to obtain some of this knowledge automatically. For example, when looking for information available on the World Wide Web, a web page containing desired information may include date of creation, last update, or details of any costs, which give some inclination on these measures.

Automatic evaluation of data quality within some databases may also be possible, for example using an integrity-checking framework such as that proposed by Caine and Embury [7]. These types of integrity checking systems produce a measure of the quality of the data sets being evaluated, resulting in a fine grained data evaluation and an overall (coarse grained) quality measure for the entire data set.

Our research has primarily concentrated on investigating the use of fine-grained quality measures, to rank data acquired from available sources. For more information on course-grained quality filtering the reader is referred to Dr Felix Naumann's work on data source ratings in [22].

4.1 User Feedback

As previously mentioned, users can be involved in the quality evaluation process by providing feedback regarding their opinions on quality of information and the sources from which it was obtained. This feedback can be used in several ways, including:

- Creating individual user profiles storing, and learning over time, their typical quality preferences;
- Creating user community quality profiles for users working together in, for example, a single research group;
- Sharing their assessment of information, and source, quality amongst a user community;
- Using feedback on the quality of the results from an information search, to assist in the development of a better search service.

The subject of multiple user feedback has been widely studied within the areas of collaborative filtering [13] and recommender systems [26], where feedback from multiple users is employed to help others find information that may meet their needs. This feedback is recorded and used to recommend similar items, such as books[1], or web sites [14], based on the feedback of previous users with similar requirements to the current user. As the idea of recommending items based upon qualities is similar in concept, the techniques developed in these research areas could be used to help obtain and share user feedback with regard to items (information, services, tangible goods) and suppliers, whether that feedback be explicit or implicit.

4.2 Static and Dynamic Quality Metrics

As can be seen above in section 3.1, a description of temporality is recorded for each quality metric; with each being marked as either *static* or *dynamic*. This information is aimed to help with the storage of historical data about acquired information, such as the price of a particular service. If a quality is marked as static then historical data about its quality can potentially be used for some time after that data is first recorded. This however is not the case for dynamic metrics. In this situation the data is likely to change frequently, meaning historical information will not be as reliable, so new information about this quality should be obtained regularly. By noting how often information on certain quality metrics changes, we can eliminate the need to constantly look for new quality values, as this is only a concern when considering dynamic metrics.

4.3 Utilising Feedback

Acquired feedback can be used as a method of adapting the various quality frameworks. This can be done either explicitly, by requesting specific attributes to be included, or implicitly by monitoring how the frameworks are used and altering them accordingly. For example, if some quality attributes are never used, over time they may be eliminated from that framework. This can be achieved by utilising ideas from research fields such as user profiling, collaborative filtering, and recommender systems.

The feedback obtained regarding which metrics are most frequently used, and by which types of user, could also be shared amongst other users, however because of user privacy considerations it is easier to share information amongst closely knit user communities (e.g., project groups) than users in general (e.g. all users with a common interest).

In our current experiments we have concentrated on looking at how the differing user quality requirements affect the results obtained from an information search, without using pre-existing ratings of data quality. As the data we are using within our current experiments is from two reputable sources no rating of the quality of this data was needed. However, quality ratings at this level will be employed during the next stage of our research, when we look into combining multiple data sources of varying quality within a single information search.

5 EXPERIMENTAL APPLICATION

To demonstrate how quality metrics can be used within an IR application two experimental systems have been created.

The first system shows how the previously discussed frameworks of quality can be created, and maintained, via a quality toolkit application. Once created, these quality frameworks can then be used within a quality-driven search system, as illustrated within our second experimental system. This second application demonstrates how these frameworks of quality can be used within a simple search system, to help the user find the information they desire from within a given dataset; in this case information regarding universities within the UK.

This chosen subject domain consists of data from two independent sources: The Times [28] and The Guardian [16]. These highly regarded UK newspapers were chosen to provide the data for this domain due to their accessibility (they are available to all via the Internet), and their information on this domain being produced independently from any university.

Each source provides data on all UK Universities, but with two important differences:

- Not all universities are included within both data sets.
- Data on the same topic is sometimes calculated differently within each source (for example, both sources provide information on research quality, but is calculated in different ways).

This provides us with a richer dataset than would be obtained if using just a single data source.

5.1 The Quality Toolkit

Before quality metrics can be used within a search application, information about these metrics must first be captured: mainly regarding how metrics relate to one another, by representing them within a hierarchical framework. It is for this reason the Quality Toolkit application was developed, as an important aspect of this project. Using previous frameworks (e.g., the original generic framework) the quality toolkit can be used to develop domain-specific frameworks by choosing relevant quality metrics, adding new metrics and, if needed, creating new quality dimensions.

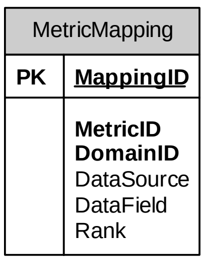
Mapping information is also inserted at this stage – the mapping of quality metrics to the real world data they represent. For example, in the university domain the quality metric of 'teaching' is comprised of five data items from within The Times [28] dataset. This mapping can be seen in Table 2, along with the rank of importance that is used when calculating values for that metric. As more data sources are incorporated into the system this mapping is liable to change, with new data items and ranks added for the newly available data.

Provider	Data Item	Rank
The Times	Teaching Quality score	1
The Times	Percentage of students completing in good time	2

Provider	Data Item	Rank
The Times	Percentage of students with good final degree	3
The Times	Percentage of graduates with good final destinations	4
The Times	Staff/Student ratio	5

Table 2: Mapping for the Teaching Quality metric

Using the Quality Toolkit this mapping information can be stored within the local mapping database, by linking the mapping information, such as that shown in Table 2, to a chosen metric in the quality framework. The storage of this mapping information is illustrated in Figure 4. Although this mapping procedure is currently a manual process a future area of research will be the identification of methods for potentially automating this process, making the addition of new data sources a faster and simpler process.



Mapping ID	ID number for quality mapping entry
MetricID	ID reference number for quality metric
DomainID	ID reference number for subject domain
DataSource	Table in which mapped data is stored
DataField	Field to be mapped to the chosen quality metric
Rank	Rank of importance of this data field when calculating the metric value

Figure 4: Storage of quality metric mapping information

Once a domain-specific framework is complete, and mapping information has been recorded, it can then be used within a separate IR application, such as our second experimental application, as discussed in the following section.

5.2 Quality-Driven Information Searching

To demonstrate how our hierarchical quality frameworks can be used in practice, an experimental search system environment has been created which allows the selection of a specific domain of interest, and presents the user with the corresponding set of relevant quality metrics from which they can select those most appropriate to their current quality need. The user can then apply weighting values to each of the chosen metrics, stating the importance of each metric, then view the best results according to those requirements from the currently available data.

Figure 5 shows a demonstration of this search system when used with data from our experimental subject domain of UK Universities, during the process of quality metric selection.

When selecting metrics for use within an information search both the individual metrics and dimensions can be selected, depending on the choice of the user. When a dimension is selected this effectively selects all metrics within it, but without the need for the user to weight each metric individually.

The option also exists for selecting a default set of metrics. This can be used to increase the speed of the search process, as the user does not then need to select all metrics individually, but rather use those that are deemed to be the most popular. These default metrics can then be edited (removing undesired metrics and including others), and then weighted in the usual way.

5.2.1 Weighting of Quality Metrics

Once a set of quality metrics has been chosen, the next stage is for the user to state weighting values for each of these metrics. These weightings are used to state the importance of each metric within the current situation, which are then used by the chosen ranking algorithms to find the best possible results, based upon these weightings.

Currently under development is a facility for also weighting the data providers. For example, in the UK Universities environment both data providers are currently rated equally, but with further development the aim is to allow the user to rate which data source they prefer, and potentially allow elimination of some data providers.

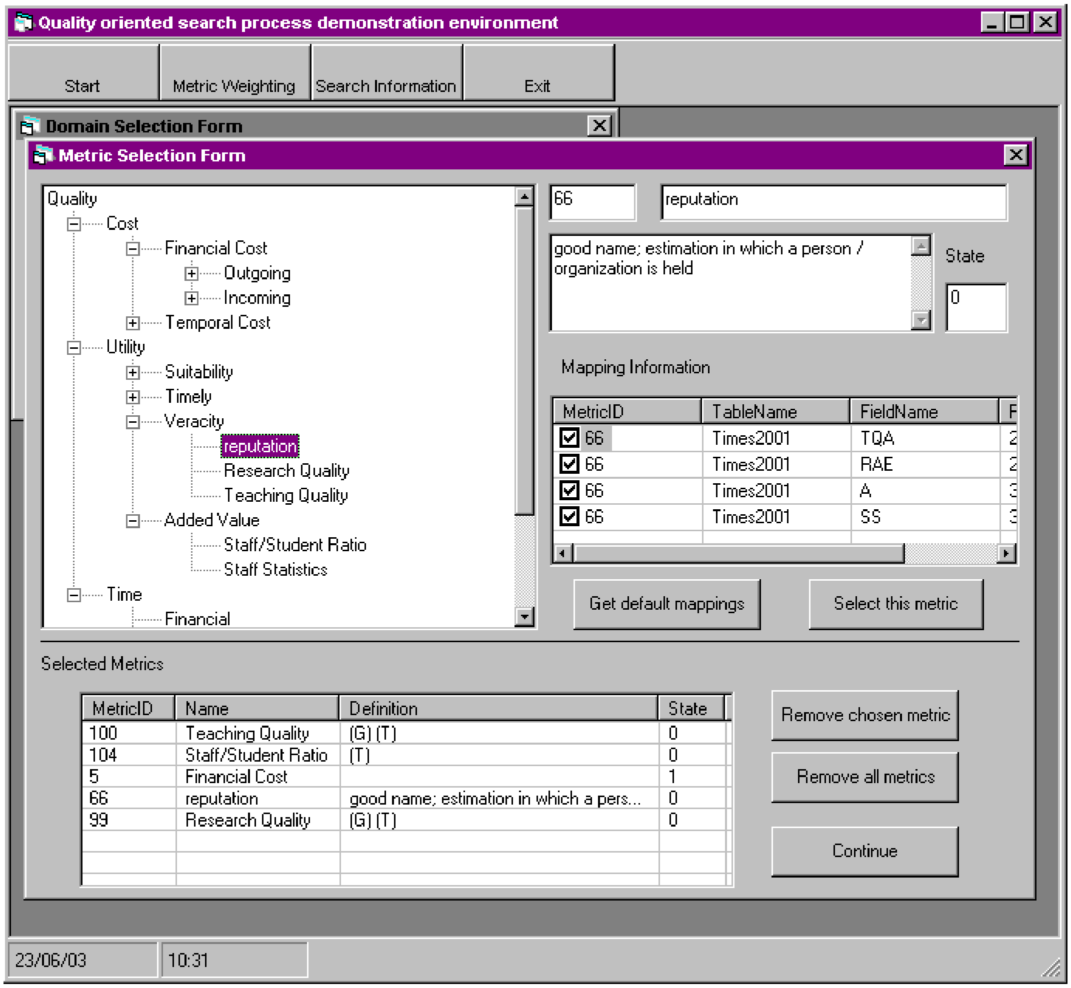


Figure 5: Quality based searching demonstration system

5.2.2 Ranking Algorithms

To find the best results based on user stated metric weightings, from the available data, several ranking algorithms have been employed.

The principal ranking algorithm used within our system is TOPSIS – Technique for Order Preference by Similarity to Ideal Solution. This algorithm is based on the principal that best result "should have the shortest distance from the ideal solution and the farthest from the negative-ideal solution." [17]. This method will therefore rank highly the results that are closest to the best possible solution available within the chosen dataset.

Other algorithms, such as the Simple Additive Weightings (SAW) ranking method have also been implemented during the development of our quality-driven search system. This allows experimentation not only on the results obtained via metric selection and weighting, but also via the different methods of calculating the best results. However, as TOPSIS has been the main method used the results presented below are those generated when using that ranking algorithm.

5.2.3 Searching for Results

To search for the best results within the available datasets the chosen ranking algorithm first calculates values for all selected metrics, using the mapping information provided, as discussed in section 5.1. When conducting multiple attribute comparisons the problem of comparing values on different scales arises. For example, it is not possible to directly compare two values when one is on a scale of 1 to 10 and another on 1 to 100. To resolve this problem, and thus enable cross-metric comparisons, a normalisation formula is applied to all calculated metric values, consequently placing them all on the 0 to 1 scale. These normalised values are then used to calculate weighted values, using the user-specified importance ratings that have also been normalised to be on the 0 to 1 scale.

The resultant values are then used by the chosen ranking algorithm to find the best results within the available datasets. These are presented to the user as an ordered set of results, with those having the highest quality values (and hence the highest ranks) being those that best meet the users' stated quality requirements.

5.2.4 Provisional Results

As can be seen from the above discussions, it is possible to incorporate quality metrics within an information search to assist a user in finding information they desire, from within available datasets. However, what is not shown is whether the inclusion of quality metrics makes any difference in the results that are presented to the user. Previous work has shown that although one quality metric has little effect, incorporating multiple quality metrics within a search significantly improves search effectiveness [32]. Our experiments have therefore focused on the hypothesis that changing quality metric weighting values will alter the result set returned to the user.

When using the TOPSIS ranking algorithm each field in the domain, i.e. each university, is given a value based on the available data for each selected quality metric, and the user-applied importance weightings. However, the exact calculated values are not of real interest within our system, as the main focus is on the order in which the results are presented to the user – with the best result being displayed at the top of the result list. We are therefore concerned only with the ranking order in which the results appear.

With this in mind, the Wilcoxon signed ranks test for statistical significance was chosen as a means to evaluate the result sets returned from a group of experiments. Although limited in its applicability within this experimental domain, this test is useful for comparing ranked results, when used with related datasets. We therefore used this test with related sets of quality metrics, when applied within the UK university domain, to assess the significance of the difference between the rankings of universities as metric values are altered.

The results of our initial statistical analysis can be seen below in Table 3. This table shows the quality metrics that were chosen for each experiment, the weightings that were applied to these metrics, and a statement as to the level of significance of the differences in result ranking order according to the Wilcoxon signed ranks test.

The first result agrees with the results obtained by Zhu and Gauch [32], in confirming that the use of one metric alone does not have a significant effect on the returned result set. This is because when ranking based upon the importance of one quality metric the actual weighting value is of little importance, as the results will always be in the same order. The only effect of changing a single metric's weighting is to change the calculated quality values for all data fields, by identical amounts, therefore leaving the final rankings unchanged.

This however is only relevant when discussing applying weighting values of a metric alone. If combined with stating a desired value for a chosen quality metric it is foreseeable that the rankings would indeed change, as the importance weighting would then be on the requirement for the highest ranking data to closely match that stated value. Within this paper we do not discuss the inclusion of a desired value factor, as that subject is the focus of future experimentation.

Test	**Metric set 1**		**Metric set 2**		**Significance Level**
	Metric	**Weighting**	**Metric**	**Weighting**	
1) One metric, with different weighting	Reputation	10%	Reputation	90%	not significant
2) Using one then two metrics	Teaching quality	75%	Teaching quality Research quality	75% 75%	0.2% = Very highly significant
3) Two metrics, with different weightings	Teaching quality Research quality	95% 5%	Teaching quality Research quality	5% 95%	0.2% = Very highly significant
4) Set of four metrics, with different weightings	Research quality Teaching quality Student completion Financial stats	50% 50% 50% 50%	Research quality Teaching quality Student completion Financial stats	90% 70% 50% 30%	0.2% = Very highly significant

Table 3: Initial Wilcoxon signed ranks significance test results

As can be seen in Table 3, the rankings of results obtained when changing weighting values of quality metrics are statistically significantly different. The results with the highest rankings, and therefore presented at the top of the returned results set, are those that are 'best' according to the chosen set of quality metrics.

Although these initial results give us a valuable insight into the usefulness of incorporating weighted quality metrics within an information search, further research is still required to fully understand and exploit this approach.

6 FUTURE WORK

As illustrated in section 5, further work is required in this area before we can be entirely confident that using quality metrics within an information search is worthwhile.

The following points are those that are to be investigated during the next stage of this research project, as discussed earlier within this paper:

- Further statistical analysis of current experimental results;

- Weighting of data source when using information from multiple providers;
- Creation of default quality metrics, for automatic selection of the most popular metrics;
- Creating more domain-specific quality frameworks, for further experimentation; and
- Searching for information within larger datasets.

The results obtained from the current quality-driven search are ranked according to quality metric weighting values alone. A future development will involve expanding this feature to also take into account metric values. Weighting values will then relate to the importance of each quality metric being as close as possible to the stated metric value.

The current user interface for this system presents the user with all relevant quality metrics for the chosen domain. This means giving the user total control over the metrics they wish to use. However, when searching for information within relatively small data sets the time taken to select all required metrics is likely to outweigh its benefits, meaning the metric selection facility is of no real value. An important stage of further development therefore consists of allowing default metric selections, with each metric having a predefined weighting value. This will free the user from having to go through the entire metric selection process, if all they desire is a quick and dirty search. The option of changing these metrics and their weightings should still exist, but will be optional.

As well as these advances there are many other directions in which this work could be taken, with sufficient time and resources. A selection of potential areas for development is as follows:

- Automatic mapping of quality metrics to available data;
- Investigating the use of user profiling techniques to automatically create quality profiles for user communities;
- Using recommender system techniques to incorporate feedback about data and source quality that can be used to refine current quality values;
- Automatic extraction of up-to-date information from dynamic data sources, such as the Internet, for use within a quality-driven information search; and
- Using personal software agents to learn user profiles and search for information within a distributed environment on the users behalf.

7 CONCLUSION

Although quality of information is extremely important to many people, little work has been done in the area of developing a framework that can be used for evaluating information quality from a user perspective, across a wide range of information domains, such as typically the ubiquitous Internet. Through our research we have investigated this problem and suggest a solution based on both a generic framework of quality, and a set of complementary domain-specific frameworks.

These quality frameworks can be used to help the consumer search for information within their chosen domain of interest, providing a framework currently exists for that domain. Although the number of available domains is currently somewhat limited, the Quality Toolkit application aids the fast creation of these frameworks. Once created they can then be reused, and shared amongst other users.

When a domain of interest has been decided upon the user can then select their desired set of metrics from the domain-specific framework of quality, and rate their importance by stating weighting values for each chosen metric. These metrics and their weights are then used when searching though available data to rank the final set of results based upon the individual quality requirements of that user. The results best meeting the users information need will therefore appear at the top of the result set.

Initial experimentation has concentrated on evaluating the changes in result ranking that occur when the weighting values of quality metrics are changed, to account for different individuals needs. Although altering the weighting value of a single selected metric has no effect on the ranking order of the results, when using multiple metrics the changes in the final ranking order are statistically significantly different. This shows that when choosing a set of metrics, personalised importance weightings can have a significant effect on the order in which the final result set is presented to the user.

Although many challenges within the field of information quality still remain [27], the work presented in this paper provides one method of using the quality of information to assist in information retrieval. The creation of a set of hierarchical quality frameworks, that can be applied quickly and easily to specific domains, is one step on the path to helping users find what they want, while combating the problem of information overload.

ACKNOWLEDGEMENTS

We would like to thank EPSRC [10] for supporting M. Burgess with a studentship (award number 99302270), and BT [5] for providing supplementary funding. Thanks also go to Mrs Wendy Ivins, Cardiff University, for providing additional support.

REFERENCES

[1] Amazon.co.uk. http://www.amazon.co.uk (Accessed on 23-6-2003).

[2] Bessell, T. L., Anderson, J. N., Sansom, L. N., and Hiller, J. E. "Surfing, self-medication and safety: buying non-prescription and complementary medicines via the internet," *Quality and Safety in Health Care*, vol. 12, no. 2, 2003. pp. 88-92.

[3] Boehm, B. W., Brown, J. R., and Lipow, M. "Quantitative Evaluation of Software Quality." *Proceedings of the 2nd International Conference on Software Engineering* , San Francisco, CA, USA 1976. pp. 592-605.

[4] Bouch, A. "Quality is in the Eye of the Beholder: Meeting Users' Requirements for internet Quality of Service." *Proceedings of the CHI 2000 conference on Human factors in computing systems* (CHI 2000), The Hague, Netherlands 2000. pp. 297-304.

[5] BT. http://www.bt.com (Accessed on 1-5-2003).

[6] Burgess, M. S. E., Gray, W. A., and Fiddian, N. J. "Establishing a Taxonomy of Quality for Use in Information Filtering." *Proceedings of the 19th British National Conference on Databases* (BNCOD 19), Sheffield, UK Springer, LNCS 2405, 2002. pp. 103-113.

[7] Caine, N. and Embury, S. M. "LOIS: the "Lights Out" Integrity Subsystem." *Proceedings of the 18th British National Conference on Databases* (BNCOD 18), Chilton, UK Springer, LNCS 2097, 2001. pp. 57-76.

[8] Cavano, J. P. and McCall, J. A. "A Framework for the Measurement of Software Quality." *Functional and Performance Issues – Proceedings of the Software Quality Assurance Workshop* , San Diego, CA 1978. pp. 133-139.

[9] CIA World Fact Book. http://www.cia.gov/cia/publications/factbook/ (Accessed on 27-6-2003).

[10] EPSRC. http://www.epsrc.ac.uk (Accessed on 1-5-2003).

[11] Firquin, B. L. "Quality is Free… But How Do You Implement Total Quality And Restructure Information Technology At The Same Time?" *Proceedings of the 20th {SIGUCCS} User Services Conference* , New York, NY, USA ACM Press, 1992. pp. 93-96.

[12] Forgionne, G. A. and Kohli, R. "A Multiple Criteria Assessment of Decision Technology System Journal Quality," *Information and Management*, vol. 382001. pp. 421-435.

[13] Goldberg, D., Nichols, D., Oki, B. M., and Terry, D. "Using Collaborative Filtering to Weave an Information Tapestry," *Communications of the ACM*, vol. 35, no. 12, 1992. pp. 61-70.

[14] Google. http://www.google.com (Accessed on 23-6-2003).

[15] Graham, L. and Metaxas, P. T. "'Of Course It's True; I Saw It On The Internet!': Critical Thinking in the Internet Era," *Communications of the ACM*, vol. 46, no. 5, 2003. pp. 71-75.

[16] Guardian, The. http://www.guardian.co.uk (Accessed on 4-6-2003).

[17] Hwang, C. L. and Yoon, K. *Multiple Attribute Decision Making: Methods and Applications* 1981.

[18] Juran, J. M. and Godfrey, A. B. *Juran's Quality Handbook (Fifth Edition)*, McGraw Hill, US, 1999.

[19] Kim, H. M. and Fox, M. S. "Formal Models of Quality and ISO 9000 Compliance: an information systems approach." *Proceedings of the 48th Annual Quality Congress*, Las Vegas, NV, Milwaukee, WI 1994. pp. 17-24.

[20] Kotz, D. and Gray, R. S. "Mobile Agents and the Future of the Internet," *ACM Operating Systems Review*, vol. 33, no. 3, 1999. pp. 7-13.

[21] Maes, P. "Agents that Reduce Work and Information Overload," in Maybury, M. T. and Wahlster, W. (eds.) *Readings in Intelligent User Interfaces* Morgan Kaufman, San Fransisco, California, USA,1998.

[22] Naumann, F. *Quality-Driven Query Answering for Integrated Information Systems*Springer-Verlag, 2002.

[23] Rangarajan, K., Swaminathen, N., Hegde, V., and Jacob, J. "Product Quality Framework: A Vehicle for Focusing on Product Quality Goals," *ACM SIGSOFT Software Engineering Notes*, vol. 26, no. 4, 2001. pp. 77-82.

[24] Redman, T. *Data Quality for the Information Age*Artech House, MA, USA., 1996.

[25] Research Project Web Site. http://www.cs.cf.ac.uk/user/M.Burgess/phd/index.html (Accessed on 9-6-2003).

[26] Resnik, P. and Varian, H. R. "Recommender Systems," *Communications of the ACM*, vol. 40, no. 3, Mar.1997. pp. 56-58.

[27] Strong D.M., Lee, Y. W., and Wang, R. Y. "10 Potholes in the Road to Information Quality," *IEEE Computer*, vol. 30, no. 8, Aug.1997. pp. 38-46.

[28] Times, The. http://www.timesonline.co.uk/ (Accessed on 4-6-2003).

[29] Wang, R. Y., Madnick, S. E., and Lee, Y. W. *Data Quality*, Kluwer Academic Publishers, 2001.

[30] Wang, R. Y. and Strong, D. M. "Beyond Accuracy: What Data Quality Means to Data Consumers," *Journal of Management Information Systems*, vol. 12, no. 4, 1996. pp. 5-34.

[31] Which? Online. http://www.which.net/ (Accessed on 4-6-2003).

[32] Zhu, X. and Gauch, S. "Incorporating Quality Metrics in Centralized/Distributed Information Retrieval on the World Wide Web." *Proceedings of the 23rd Annual International ACM SIGIR conference on Research and Development in Information Retrieval* (SIGIR-00), Athens, Greece 2000. pp. 288-295.