Context for Semantic Metadata

Kenneth Haase
beingmeta, inc. & Media Lab Europe
68 Bailey Street
Boston, MA 02124
kh@beingmeta.com

ABSTRACT
This article argues for the growing importance of quality metadata and the equation of that quality with precision and semantic grounding. Such semantic grounding requires metadata that derives from intentional human intervention as well as mechanistic measurement of content media. In both cases, one chief problem in the automatic generation of semantic metadata is ambiguity leading to the overgeneration of inaccurate annotations. We look at a particular richly annotated image collection to show how context dramatically reduces the problem of ambiguity over this particular corpus. In particular, we consider both the abstract measurement of "contextual ambiguity" over the collection and the application of a particular disambiguation algorithm to synthesized keyword searches across the selection.

Categories and Subject Descriptors
H.2.1 Logical Design, H.2.3 Data Description languages, H.2.4 Multimedia databases, H.2.7 Database Administration: *Data dictionary/directory,* H.3.3 Information Search and Retrieval, H.3.7 Digital Libraries, I.2.7 Natural Language Processing, K.1 The Computer Industry

General Terms
Management, Measurement, Performance, Economics, Algorithms.

Keywords
Metadata, information retrieval, context, disambiguation, multimedia databases.

1. The Value of Metadata
Detailed and precise metadata will be the key to the next generation of applications that will realize the potential of the new digital media. Especially with largely opaque media (image, video, audio), metadata provides the handles by which programs can search, arrange, and repurpose the digital media that are quickly becoming ubiquitous. However, the generation of that metadata and the economics of that production and application remain problematic.

MM'04, October 10–16, 2004, New York, New York, USA.

We believe that the economics of metadata are subject to a principle analogous to Metcalfe's law for the economics of network technologies [5][12], in particular that

> "the value of metadata rises as the product of the log of the corpus size and the log of the size of the user community"

For example, good metadata would not be a crucial issue for small databases (100s or 1000s of items) or a handful of users, since a simple organizational scheme (based on time, place, keywords, etc.) could be combined with personal knowledge to allow fast, relatively reliable, retrieval and identification of relevant images. However, as the total number of images grows or the community size increases, the value of metadata increases substantially.

Applying this principle to the growing pool of digital media and clear market trends (omnipresent cameras, huge disks, ubiquitous broadband), suggests an oncoming transition in the traditional economics of data and metadata.

2. The Metadata Twist
While the economic significance of metadata increases with the size of the available content, the average economic value of the content must, at the same time, decrease with the amount of available content. This pair of trends leads to a projection that we call the "Metadata Twist", illustrated in Figure 1.

As media technologies improve and spread, there will be a gradual transformation where metadata will become more valuable (on average) than the content it describes. This counterintuitive twist arises as the human resource of time and attention remains fixed while the pool of accessible media increases at least exponentially.

The potential in this transformation is enormous and motivates the investment in technologies and capabilities for dealing with the metadata that will become increasingly valuable. Considered carefully, it also focuses attention on the importance of high-quality **precise** metadata.

3. Quality, Precision, and Semantic Metadata
Not all metadata are created equal. One useful definition of metadata is "any data which conveys knowledge about an item without requiring examination of the item itself." Because metadata derives its value from saving human time and attention, it must be effective at distinguishing relevant and irrelevant or redundant content. For automatic storytelling, quality metadata is even more important, since a given

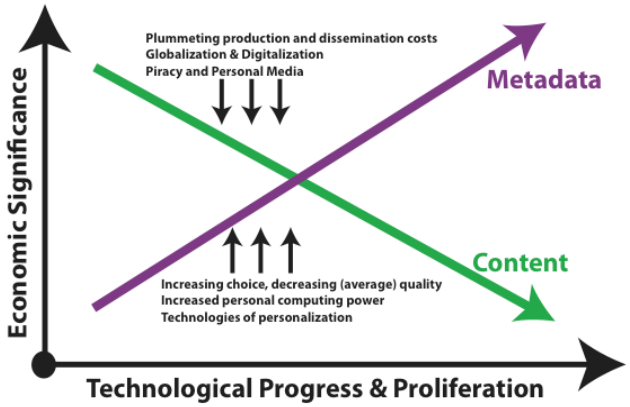

Figure 1: The Metadata Twist

presentation may rely on complex sequences of discriminations regarding features and connections which can serve narrative goals.

One of the chief characteristics of metadata quality is **precision**: metadata needs to be precise enough to effectively distinguish between different items so that it can select the right one without requiring that a person examine the items (absorbing their precious time and attention). Precision is equally important to determine what items should be **excluded** as included, whether for reasons of irrelevance, inappropriateness, or redundancy. In either case, the more precise the metadata, the more valuable (in the terms described above) it is.

There are two significant problems with precise metadata. First, in conventional databases and metadata models, increasing precision of description reduces the overall recall performance. If you describe an image as a "German Shepard," a person searching for pictures of "Pets" will not find it. Second, precise metadata may require some human attention to produce it, the same attention we're trying desperately to save by using it in the first place.

The problem of diminished recall ("German Shepards" vs. "Pets") can be addressed in large part by using "semantic metadata" which links related terms to one another, so that "German Shepard" is connected to "Pets" in some manner. Semantic metadata differs from traditional taxonomies or structured thesauri in two important ways:

- it provides articulated patterns of reference, describing (even if only in natural language) how terms map to content; and
- it provides operational rules of inference explaining how and when terms can be expanded to other terms.

These criteria are interdependent. For instance, to insure that we can expand a term like "Fish" appropriately, we need to distinguish "Fish, the food" from "Fish, the animal". Conversely, because we may distinguish "German Shepard" from "Labrador Retriever," we need to be able to infer that they are both kinds of pets (or at least domesticated animals).

While we have focused here on terminological precision, other sorts of precision are equally important. Davis [2] makes a similar argument for stream-based, rather than clip-based, video representations (in which all annotations have time indices), for the same reasons of more exact retrieval and better automatic manipulability.

In conclusion, precise description of large collections requires semantic metadata and the application of semantic metadata has its own requirements. We consider these in the next section, especially the need for multiple schematizations or taxonomies and the need (into the foreseeable future) for human annotation of content.

The second issue, the effort required for precise annotation, is potentially more serious. The argument for precise annotation is that we are strategically shifting time and attention from the search process to the initial annotation process. If the number of searchers is large and their time is valuable, this can be a strong argument. On the other hand, when the number of documents is very large, the cost of precisely annotating each one (even ones which searchers will never see) may outweigh this advantage. However, it is important to not take naïve assumptions about the cost of precise annotation for granted. As we will discuss below, new techniques --- such as the use of context --- can dramatically reduce the cost of annotation.

4. Metadata and Purposes

Metadata, and especially semantic metadata, is necessarily a **representation** of whatever content it describes. Like any representation, it selects and summarizes, reducing the content of the media itself to a more compact and easily manipulable form. This manipulability is the raison d'etre for the metadata in the first place.

One important property of representations is that they are artifacts and especially purposive artifacts. This means that every representation has a set of associated purposes and the representation's criteria of selection and summarization reflect those purposes. For instance, the metadata associated with photographs in a news organization might be very different from the metadata associated with photographs by a fashion company or by an advertising agency.

The multiplicity of purposes leads directly to a need for multiple descriptions or, technically, multiple schematizations or

taxonomies. Because there is no organizational scheme that will satisfy all purposes, we need to allow a diversity of schemes while providing ways to heuristically connect among them. Diehard positivists might argue that every organizational scheme might bottom out in some kind of basic language, but even if this is so (and there are strong reasons to believe otherwise), there is still a pragmatic reason to have a plurality of schemes which can compactly represent complex deeply nested expressions in the putative "universal primitive language".

It is important to distinguish broad-coverage representations (which are common) from general-purpose representations (which we are arguing against). Metadata standards such as the Dublin Core are broad-coverage but special purpose representations. Because they are designed to serve very general, typically bibliographic purposes, they can apply generally to a broad range of media items. Likewise, broad-coverage annotation schemes like Media Streams [2] focus on visual description for particular purposes such as reusing content for particular kinds of storytelling.

In addition to requiring a multiplicity of descriptive schemes, precise semantic metadata eventually requires some level of human annotation and involvement. The reason is that purposes are essentially socio-technological constructs that reflect what people do, what people need, and what technology enables them to do. This complexity is what Hofstadter calls "AI-complete": it cannot be solved without solving the problem of creating intelligent machines, which does not constitute a near-term solution. While most activity in annotation has dwelt on automatic categorization (by, for instance, image processing), progress has been relatively slow even at the lowest level of reliably recognizing settings or individuals.

As we indicated above, using people as a source of precise metadata is a strategic shift of time and attention from actual and potential users to the people producing the metadata. The details of this shift will depend on the nature and needs of those users, the economic model for the entire system, and the skills and tools of the metadata producers. However, if we are requiring precise metadata (which we've argued will become more and more important), we need human sources of metadata and it is useful to think about how to generate this metadata as effectively and inexpensively as possible. In order to understand how to do this, we need to look at the processes that generate both automatic and manual metadata.

5. Sources of Metadata

The process of generating accurate and useful metadata can be visualized as a flow that generates features from sources in the external world. Crucially, these features are **not independent**, leading to the importance of **context**: some features consistently co-occur with each other and this provides an important potential constraint on the processes that generate metadata. Figure 2 shows this process, which we call the **metadata pipeline**. We begin by distinguishing two original sources of metadata: measurement and intervention.

Measurement involves the mechanical extraction of features from media content or context. It can include analysis of pixels or motion in an image, the time or location of a capture event (say with GPS coordinates), or sophisticated extraction of higher-level features of the raw signal.

Intervention requires action of a human being with respect to the content that does not change the content itself. It can include the assignment of a keyword, the classification into a folder, or the juxtaposition of content items together in a presentation.

The distinction is helpful because metadata arising from intervention, when interpreted correctly, tends to more reliably reflect the human purposes upon which representations are naturally based. For example, the Google search engine focuses on features such as patterns of web linking and URL text (i.e., domain names, file paths and names, etc.), in judging the similarity of documents to queries. These features all involve choices made by people in creating links, registering domain names, choosing file names, and organizing directory structures.

Starting from raw measurements or interventions, the metadata process attempts to derive useful metadata: representations that capture actual or potential human purposes. Often, the raw datum of measurement or intervention is not refined enough to use as metadata to drive searching or reusing media. For example, an assigned keyword may be ambiguous with respect to potential users' "information needs" or a media item may be in audio but users will be searching using text. In each of these cases, processing goes from the raw material of measurements and interventions to more refined purpose-linked features.

The distinction between measurement and intervention in origin can be coupled to a distinction between **identification** and **interpretation** in processing. Identification extracts features from measurements; interpretation extracts features from interventions. Both of these processes are necessarily imprecise and heuristic since they are making guesses from partial information. Interpretation is making guesses about the past: "They meant fish as animals" or "they didn't like this clip". Conversely, for media reuse in particular, metadata identification is making guesses about the future: "people will see George W. Bush in this cartoon" or "this setting is recognizably the same as this other one".

This imprecision is the source of the "semantic gap" [3][16] since measurements, at least in isolation, underspecify the semantic description of contents or purposes. In order to close this gap, both identification and interpretation can draw on outside knowledge of various sorts, whether feature templates for object recognition, mappings of ambiguous keywords to unambiguous concepts, or geographic databases that identify places from measured coordinates.

These processes can fail in two different ways: they can fail to identify or interpret an actual feature or they can identify or interpret features that do not actually apply. For the rest of this paper, we will focus on this second case: the misattribution of features that do not actually apply.

Before we move on to this focus, it may be helpful to briefly unpack the messy bundle tied up in the words "actually apply" above. The acid test for the "accuracy" of metadata is whether it works in the processes of selection or generation in which it is enlisted. In search, if relevant items are found and irrelevant or redundant items discarded, then we can say that the metadata was correct. In repurposing or automatic story generation, if the generated presentation fulfilled its purpose (whether education, entertainment, seduction, or convincement), the applied metadata was correct.

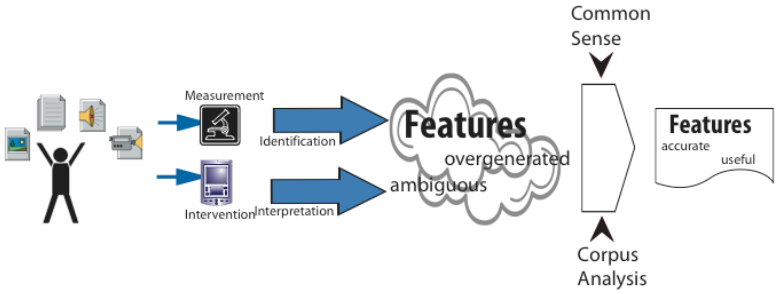

Figure 2. The Metadata Pipeline

This definition is unapologetically pragmatic but it also challenges a common practice in evaluating metadata. Human beings, in the aggregate, are often taken as a "gold standard" for metadata, permitting the discounting of distinctions that humans do not reliably make. However, the use of metadata is typically individual and aggregate generalization may be a poor guide for such applications. Just because the average individual may not be able to distinguish (at least by name) oaks from dogwoods, doesn't mean that metadata need not. Or more personally, if someone is looking for old family pictures of their mother, photographs of her twin sister just won't do.

This particular case of the "twin aunt" highlights the ambiguity inherent in metadata generation *regardless of the sophistication of its signal processing*. In turn, this ambiguity leads to the problems of misattribution of metadata. One way to minimize this misattribution is to expand our scope from individual features to the constellation of features that define the "context" in which the individual features are identified or interpreted.

6. Overgeneration and Ambiguity

The metadata production process produces possible features by either interpretation of human interventions or measurements of content characteristics. One of the ways in which this process fails is when it generates features that do not actually apply to the content. This overgeneration is a form of ambiguity: a measurement or intervention has several possible identifications or interpretations and all of them are produced.

One of the ways to resolve ambiguity is to combine what we know reliably about an item with external knowledge that tells which features are likely or unlikely and to discard the ones that are unlikely. We refer to the "known facts" as the **context** in which the disambiguation of a given possible feature is performed. This context, combined with background knowledge, can help in disambiguation.

For example, if we know that we are at a wedding, we can guess that the keyword "groom" is unlikely to refer to a stableboy and that the keyword "cake" is likely to refer to the food and not the soap. This external knowledge approach is taken, for instance, in Aria [10] to identify relevant concepts and then expand searches as in an **inference network** [17].

This kind of external knowledge can come from a variety of sources. One source is common sense knowledge bases, such as CYC [9] or ConceptNet [11]. Another, more direct source, would be direct access annotated corpora. We will use a variation on this latter source of knowledge to determine the possible role that context could play in disambiguation of keywords to disambiguated concepts. We will then evaluate a simple algorithm to use the corpus itself as a source of background knowledge for disambiguating synthesized ambiguous annotations.

7. Ambiguity in Context

To look at the role that context may be able to play in some sorts of disambiguation, we will be considering the characteristics of image collections annotated with concepts. In particular, we will be comparing the absolute ambiguity of terms (the meanings they may have) and their relative ambiguity (how many meanings they have when known to be associated with other concepts). What we are attempting to measure here is, essentially, how much the interdependence of terms (which makes context significant) can reduce the semantic gap.

In order to formalize this definition, we consider a set of media items I annotated with concepts drawn from a vocabulary C, each of which may be ambiguously referred to by natural language terms T. Given these domains, we introduce several functions:

FindContent(concept) maps concepts into media items;

GetConcepts(item) maps media items into concepts;

FindConcepts(term) maps terms into concepts;

GetTerms(concept) maps concepts into terms.

The absolute ambiguity of a term t is thus:

$$\left|FindConcepts(t)\right|$$

The relative ambiguity of a term t with respect to a subcollection $J \subseteq I$ would then be:

$$\left|(FindConcepts(t)) \bigcap (\bigcup_{j \in J} GetConcepts(j))\right|$$

Given these definitions, we define the contextual ambiguity of a term t given a concept X as the relative ambiguity of t with respect to the corpus of items annotated with both a meaning of t and the concept X, e.g.

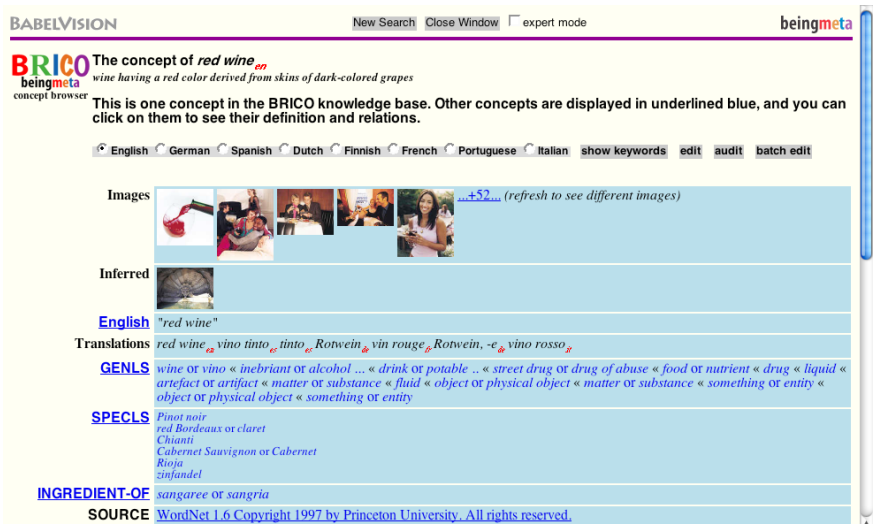

Figure 3. The description of a BRICO concept

$$Ambig_X(t) = \left| (FindConcepts(t)) \bigcap (\bigcup_{j \in FindContent(X)} GetConcepts(j)) \right|$$

This indicates the ambiguity of t applied to a media item if we already know that the concept X applies to it. We can use this to define the mean ambiguity of t that is the average ambiguity over all concepts that co-occur with possible meanings of t. Note that the mean ambiguity is a pessimistic metric: it is the expected value of the ambiguity given a single random contextual element. In practice, we are likely to have multiple contextual elements which will not be random.

We will consider these metrics applied to a relatively large (75,000 image) image database annotated with disambiguated concepts from the BRICO [6] knowledge base. The image database has been provided by Digital Vision Online (DVO), an international royalty-free provider of stock photographs. The BRICO database was generated from the amalgamation of multiple sources, most significantly the WordNet online lexical thesaurus [4][13][14]. A sample concept from BRICO is shown in Figure 3.

BRICO extends WordNet with more concepts, additional linkages, limited inference capabilities, and links into other languages. Concepts in BRICO represent disambiguated word meanings and are linked to other concepts through a variety of relationships.

The DVO collection was used in previous work [7][8] to explore the possibility of non-expert conceptual annotation. A sample image from the collection, together with its annotations, is show in Figure 4. For this study, we used expert keywords assigned by DVO and hand audited and automatic mapping from those keywords to concepts in BRICO. The DVO keywords are substantially monosemous by design, meaning that each

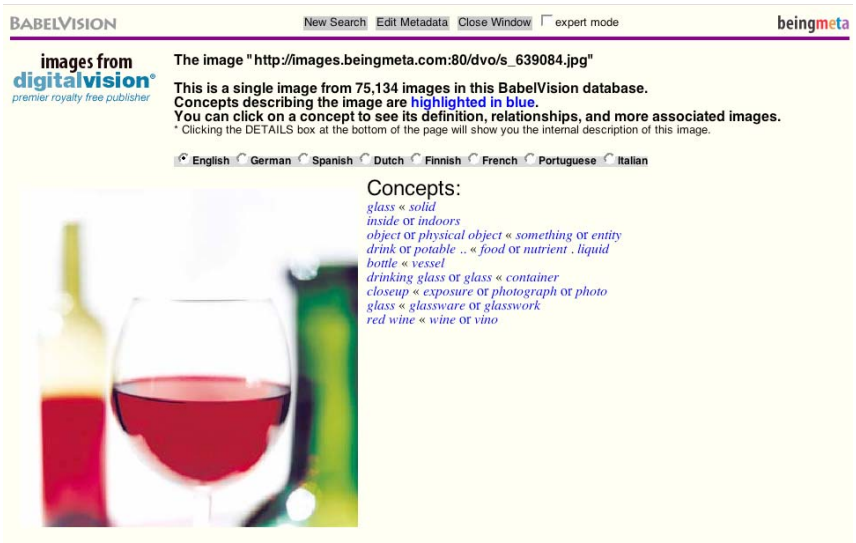

Figure 4. An image and its annotations

keyword can be mapped to one concept in BRICO or, occasionally, collections of related concepts (such as the noun "investment" and the verb "invest").

The resulting database is heavily annotated, averaging 12 concepts per image. The annotated corpus uses 3,759 concepts overall which encompass 6,199 English words (BRICO also includes foreign language words for concepts). 6,016 of those 6,199 words map unambiguously into single concepts. The remaining 183 words map into multiple meanings (between 2 and 4) averaging 2.1 meanings per word.

8. Contextual Ambiguity

The contextual ambiguity for each of the 183 ambiguous terms ranges from 1 to 2.4, averaging 1.2 meanings for each term. This measures the degree, over the entire corpus, to which a given bit of context will disambiguate the meaning of the term.

The average of 1.2 meanings above reflects the fact that some concepts may not provide any help in disambiguating a term. A more revealing metric is to look at how many of the possible contextual cues perfectly disambiguate the term. For example, the term "bag" has 4 possible meanings over the corpus and these meanings co-occur with 213 other concepts. 142 of these concepts (67%) perfectly disambiguate the term "bag" to its correct meaning. Over all the ambiguous terms, the percentage of perfect disambiguators ranged from 0-100%, averaging 78%.

9. A Simple Algorithm

The metrics above suggest that context can provide a strong constraint on keyword disambiguation. In order to examine this further, we can implement a simple algorithm to automatically perform such disambiguation.

The algorithm starts with a set of keywords $K \subset T$ and scores each item $j \in I$ based on the potential applicability of the terms $k \in K$ to the item. This applicability is ambiguous: a term k applies to j if any of the meanings of k are associated with the item j. Thus, the score of an item j would be the size of the overlap:

$$\left| (GetConcepts(j)) \bigcap (\bigcup_{k \in K} FindConcepts(k)) \right|$$

Then, for each keyword k, the algorithm finds the highest scored item(s) to which k applies and picks the meaning(s) of k associated with those items.

We can assess this algorithm by picking random items from the database and generating putative keyword sets based on the concepts describing them. We then execute the algorithm (being sure to remove the item itself from consideration) and compare the resulting disambiguations to the actual concepts.

Resulting from this comparison, extra concepts indicate **retained ambiguity** while missing concepts indicate **misinterpretation** or failure of the algorithm to find the correct meaning (miscuing from the context).

Note that this task is simpler than free text disambiguation [15] because the context of each corpus item is already disambiguated and the domain itself is mostly limited to physical description. The relative role of these two factors will determine how effectively this simple algorithm would scale to both more general (non-physical) descriptions and corpora that are unevenly annotated.

To evaluate this simple algorithm, we applied it to 10,000 queries synthesized from the DVO corpus and computed mean retained ambiguity and failure both per query and per keyword.

Per query, the average retained ambiguity was 0.12, meaning that roughly 10% of the queries could not be disambiguated completely, though the error was generally limited to only one of the keyword terms. The average misinterpretation rate was 0.05, indicating that the algorithm misinterpreted keywords in 5% of the queries. But again, this error was typically limited to a single keyword.

Per keyword, the average retained ambiguity was 0.01, indicating that only 1% of the keywords were not uniquely disambiguated. The average misinterpretation rate was 0.004, indicating that the algorithm misinterpreted only 0.4% of the keywords.

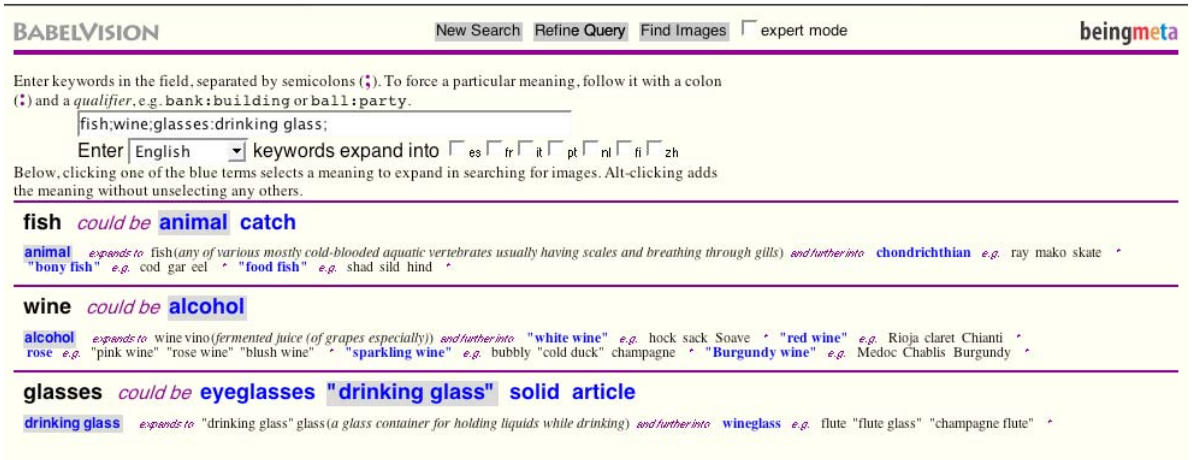

Figure 5. Disambiguating a Query

These results show that, in a richly annotated corpus, context can practically serve as a very strong constraint on disambiguation of keywords. The results are substantially better than the theoretical "mean contextual ambiguity" calculated in Section 9 (the 1.2 meanings per term) because the algorithm implicitly considers combinations of disambiguating concepts rather than single concepts. This was not surprising as we noted above that the mean contextual ambiguity was a pessimistic metric.

The application of the algorithm in practice can be seen in Figure 5, where it is used in the BabelVision interface to the DVO collection of 75.000 images. The keyword text of the query is in the box at the top of the screen. At the bottom of the screen, we see that the algorithm has selected the "drinking glass" meaning of the query term "glasses" rather than the more common (in this corpus) concept of "eyeglasses".

As we mentioned above, the scalability of this simple algorithm still needs to be determined for domains with less concrete descriptions and with corpora that are less reliably or comprehensively annotated. These are both areas for future work.

As mentioned above, the results are also favorable compared to free text meaning disambiguation because terms in each context are disambiguated and the domain is relatively limited (image description).

10. Leveraging Measurement

In the economics of metadata, there is a logical focus on measurement over intervention, even though intervention generally leads to more reliable and precise features. The reason is that the economic expense of measurement and identification is typically both lower and more easily capitalized financially. Furthermore, the technologies of measurement and identification (sensing, communication, and computation) have all been growing exponentially less expensive, enhancing this advantage.

While the purposive character of metadata ensures that intervention must remain an important source of semantic metadata, it is interesting to consider how the use of measured context might constrain interpretation. We can get a rough sense of this by repeating the analysis we did above, but limiting possible contextual cues to cues that might be easily measured. In particular, we will restrict the context considered to locational context. This includes both geographic information (this photo is in New York City) and locofunctional information (this photo is in a restaurant). Location-based technologies are becoming more pervasive and less expensive and the growth of location-based services means that this sort of information may soon readily be available at the point of capture.

We can experimentally separate out locational from non-locational concepts by using the **semantic class** that most BRICO concepts inherited from WordNet. The semantic class noun.location is used to identify both geographical and locofunctional concepts. When we repeat the analysis above, but reduce the space of "co-concepts" to these concepts, the contextual ambiguity drops to 1.15 meanings (ranging from 1 to 2.26 meanings per term). This is an improvement on the 1.26 meanings per term (ranging from 1 to 3.29 meanings per term) in the general case.

For example, consider the term "board" which has multiple meanings, including circuit boards, wooden boards, and the act of entering some for of transport. However, locational concepts, such as "house" (for wooden boards), "airport" (for entering transport), or "laboratory" (for circuit boards) decisively remove this ambiguity, making the term "board" contextually monosemous.

Significantly, the overall percentage of unique disambiguators is also higher when considering only locational concepts: 86% as opposed to 77%. These unique disambiguators, like the locational concepts listed above, make the term contextually unambiguous. The significance here is that locational categories are much more reliably identified (given infrastructure such as cellular or GPS networks) than signal-based categories but can provide a powerful constraint on disambiguating other features.

11. Future Work

More empirical study of the simple algorithm above is merited and will be ongoing over the coming year. In particular, we will look at the correlation between disambiguation performance and the size of the keyword set provided to the algorithm.

As specified, the algorithm does not take advantage of the inference possibilities in the annotation language. One obvious extension to be studied is the use of those inference relations (for instance, between general and specific concepts) to improve the performance of the algorithm, especially as smaller sets of keywords are likely to decrease the overall performance.

Another promising direction is to look at the use of monosemous references (terms with a single meaning) to bootstrap the creation of an annotated corpus and then to use that annotated corpus to disambiguate the polysemous terms.

Finally, it is important to look more closely at the interaction between identified (measurement-based) and interpreted (intervention-based) features. The cursory analysis above used locational concepts as a proxy for actual locational information. It will be interesting to work with collections which have both sources of information and see the degree to which these different sorts of features can constrain one another. Of special interest would whether this contextual approach can effectively make some unreliable identification algorithms (such as people spotting) more reliable.

12. Conclusions

We have argued that quality metadata will become increasingly important and significant as collection sizes and user communities grow. Further to this increasing importance, we have proposed that quality metadata must necessarily be precise and semantic (disambiguated and supportive of automatic inference). Looking at the problem of generating such metadata, we showed that human generated annotations of various sorts would remain extremely important in practice.

We discussed the issues in generating semantic metadata and particularly considered the problem of constraining overgeneration of semantic metadata and linked this overgeneration, in particular cases, to the problem of ambiguity. Looking at a particular richly-annotated collection of images, we found that contextual information could reduce this ambiguity dramatically.

13. References

[1] Brooks, K., Metalinear Cinematic Narrative: Theory, Process, and Tool, *MIT PhD disseration*, April 1999.

[2] Marc Davis. "Media Streams: An Iconic Visual Language for Video Representation." In: *Readings in Human-Computer Interaction: Toward the Year 2000*, eds. Ronald M. Baecker, Jonathan Grudin, William A. S. Buxton, and Saul Greenberg. 854-866. 2nd ed., San Francisco: Morgan Kaufmann Publishers, Inc., 1995.

[3] Dorai, C. and Venkatesh, S. Computational Media Aesthetics: Finding Meaning Beautiful. IEEE Multimedia, 8 (4). pp. 10-12.

[4] Fellbaum, C. ed., *WordNet, An Electronic Lexical Database*, MIT Press, Cambridge, MA (1998).

[5] Gilder, G., "Metcalfe's Law and Legacy," *Forbes ASAP*, 13 September 1993.

[6] Haase, K. Interlingual BRICO. *IBM Systems Journal 39*, 3/4 (2000).

[7] Haase, K., Guaraldi, B., and Tamés, D. SBIR Phase I Report: Non-Expert Conceptual Annotation. Report submitted to the National Science Foundation, Project Number 0232731, July 28, 2003. Available at www.beingmeta.com/pub/nonexpertannotation.pdf

[8] Haase, K. and Tames, D., Babelvision: Better Image Searching Through Shared Annotation, in *ACM Interactions*, 11(2), March/April 2004.

[9] Lenat, D.B. and Guha, R.V., *Building Large Knowledge-Based Systems: Representation and Inference in the CYCProject*," Addison-Wesley Publishing Company, Reading, MA (1990).

[10] Lieberman, H., Rosenzweig, E. & Singh, P., Aria: An Agent For Annotating And Retrieving Images, *IEEE Computer*, July 2001, pp. 57-61.

[11] Liu, H. & Singh, P., ConceptNet: A Practical Commonsense Reasoning Toolkit. *BT Technology Journal, upcoming.* Kluwer 2004.

[12] Metcalfe, R., "There Oughta Be a Law," *New York Times*, 15 July 1996.

[13] Miller, G.A., "WordNet: A1 Lexical Database for English," *Communications of the ACM* **38**, No. 11, 39–41 (1995).

[14] Miller, G.A., "WordNet: An On-Line Lexical Database," *International Journal of Lexicography* **3**, No. 4, 235–312 (1990).

[15] Resnik, P., "Word Sense Disambiguation in NLP applications", in *Word Sense Disambiguation: Algorithms and Applications*, Eneko Agirre and Philip Edmonds (eds.), Kluwer, forthcoming. Kluwer, forthcoming.

[16] Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., and Jain, R. Content-Based Image Retrieval at the End of the Early Years, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, 2000; pp. 1349-1380.

[17] Turtle, H. & Croft, W.B., Inference networks for document retrieval, *Proceedings SIGIR 1990*, ACM Press.