Combining Image Context Information

Randi Karlsen
Department of Computer Science
University of Tromsø, Norway
randi@cs.uit.no

Joan Nordbotten
Department of Information and Media Science
University of Bergen, Norway
joan@infomedia.uib.no

Abstract

Current techniques for content based image retrieval have known shortcomings that make it difficult to search for images based on their semantic content. This leads to the well-known semantic gap problem. To address this problem, we propose utilizing context information, which is available from multiple sources, such as that generated by the camera at image capture, sensor data, context sources on the Internet, and information about environments where an image is used (such as an image collection or a document). In this paper we identify different types of image context information, and describe how context data from multiple sources can be combined so that it is available to the image retrieval process. We further show how image context information can be represented using a new context descriptor vocabulary CTXT.

1 Introduction

There is currently a mismatch between user needs and current capabilities of content-based image retrieval (CBIR) systems. While information seekers typically describe image queries in terms of the semantic content in the required image, CBIR systems rely on the structural characteristics of images for image selection. This makes it difficult to search for images based on their semantic content [1].

Image context information is being used in some image retrieval systems to support the retrieval process. Currently the most used context information is date/time and location which can be used for relating an image to an event or to identify nearby objects (that might be depicted in the image). We believe that an extended use of image context information can enhance automatic identification of image semantics and thus may contribute to improving image retrieval systems.

This paper was presented at the NIK-2008 conference. For more information, see http://www.nik.no/.

In this paper we identify a variety of context information and distinguish between two types of image environments. Capture Context information describes the environment in which an image is captured. Usage Context information describes the environment(s) in which an image is used, typically image collection(s) or illustrated document(s). Thereafter, we discuss how information from multiple image contexts can be combined in order to infer semantic content of images. We present elements of a new context specification vocabulary, CTXT, and demonstrate how image context information can be represented so that the characteristics of each context type are sustained and made available for use in image retrieval.

The remainder of the paper is structured as follows. Section 2 presents a background on image retrieval and current use of image context information. Section 3 introduces two image context classes, while Section 4 describes how context information from these classes can be combined, and Section 5 describes how image context information is accumulated. In Section 6 we present and exemplify a subset of the context vocabulary CTXT. Section 7 concludes.

2 Background

Image retrieval

Image retrieval techniques can be classified according to two main approaches; text-based and content-based retrieval. In *Text-based image retrieval* (TBIR) image features are extracted from image annotations that may describe semantics of objects in the image as well as the situation or activities depicted in the image [2]. Unfortunately, annotations may be biased, since they represent the annotator's understanding of the image and are described in his/her vocabulary. In addition, annotating images is primarily a manual, time consuming task, which makes it likely that an image is annotated with only one or a small subset of its possible semantic interpretations.

Content-based image retrieval (CBIR) [3] has been developed to support image retrieval in areas where textual descriptions of images are either not possible or impractical (such as satellite images, finger prints, and x-rays). The underlying idea of CBIR is image matching, where the structural characteristics of an input image are matched to the structural characteristics of the images in the database. CBIR supports automatic registration of low-level image features, such as color, shape or texture, but lacks the support for image retrieval based on high-level semantic concepts. This semantic gap [1], which represents the mismatch between semantic user requests and the capabilities of current CBIR systems, is a major problem for general CBIR systems.

Image context

A frequently used definition of context is found in [4], where Dey describes context as: "any information that can be used to characterize the situation of an entity", and an entity as "a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and application themselves". In this paper, the principle entity is an image, and our focus is on the information characterizing the image that is considered relevant for image management (including image retrieval and presentation).

Time and location characteristics are currently the most utilized context

information for an image. Image capture time can be used to cluster images into collections, as done in [5, 6] where the objective is improved image browsing. GPS coordinates identify exactly where an image was taken, and can be used as a basis for determining relationships between the image and surrounding objects, such as buildings, areas, or natural phenomenon. A number of systems combine image location information with maps, see for instance Flickr¹, Smugmug² and Woophy³. Location-based image search is described in [7], and in [8] a user's physical location is used for navigating through an image collection.

A combination of geographical location and time may be useful for linking an image to an event, such as an annual festival, and in [9] time and location are used as a basis for detecting location names, weather information, local time, daylight status and season. Identity of the photographer together with social context is in [10] used for automatically resolving the identity of subjects in an image.

The work of [9] reports on a user study measuring how effective users believe a particular set of contextual metadata can be for image retrieval. The results show that users believe the following context information can be useful: location names, distinction between indoor and outdoor images, number of people in the image, time of day, event, season, year/month, light status, and temperature. Camera setting, color, and information about where the image is stored where considered less useful.

A user study in [11] reports that there is a user interest in supplementing images with additional data, including location, activity, environmental information (e.g. temperature and light conditions), physiological parameters (e.g. heartbeat), and information about the current social situation (e.g. surrounding people).

3 Image context classes

In this paper we distinguish between two context classes, Capture Context and Usage Context, representing different types of image environments.

- Capture Context represents information about the environment in which the image was taken, and may include context information such as date/time, location, sensor data, information about an area or objects and the purpose of the image.
- Usage Context represents information about the environment in which an image is used, for instance in an image collection or in a document where the image is used as illustration. Context information can be a textual description of a collection, a document abstract, keywords describing the collection or document, or text surrounding an image.

Capture Context can be collected from a variety of sources, from physical sensors to software systems available as services for instance on the Internet. There are many different types of Capture Context information that potentially can be added to an image, but we believe that not all context information is equally relevant to an image. The required context information will change depending on for example, the image type, location of capture, interests of the image owner, and availability of context sources.

¹http://www.flickr.com/

²http://maps.smugmug.com

³http://www.woophy.com/map

Context information can be collected in a chained manner, where already obtained information can be used as a basis for capturing new context information. For example, a context source can, based on GPS coordinates, provide a city name, while a different source use the name to obtain a description of the city. The combination {date/time, location name} can provide information about temperature, humidity and wind speed at or near the place of image capture, as well as the season or daylight status.

Usage Context information can be automatically collected from collection metadata, document metadata and/or document content. An image collection is a set of images that has one or more properties in common. Typically, the images in the collection are brought together for some specific purpose, such as recording i) a specific event, ii) a time period or, iii) thematic content [12]. Collection-level metadata may include information such as collection title, application area, purpose, and topic descriptions. Information about the collection owner can also indicate potential themes of the images in the collection. Document information useful as Usage Context information includes document title, abstract, keywords, text surrounding an image, intended audience and category of the document.

The different sources of Usage Context information provide somewhat different semantic meaning. A collection description or document abstract does not represent a direct description of any of the images in the collection/document, but can give an indication of what an image is about. Likewise, keywords describing a document or collection may capture the essence of the collection/document in a few, carefully chosen, terms, but will have a more distant meaning for a specific image in the collection/document. In contrast to document descriptions, surrounding text is more likely to describe the particular image that it surrounds, and may therefore give a more direct indication to the content of the image. This illustrates that we can attach different interpretations to the different sources of Usage Context information. It is therefore important that the representation of image context information distinguishes between the different sources.

Usage Context has only been used in previous work to a limited extent. In research such as [13, 14] and systems such as Google Image Search⁴, keywords characterizing an image are automatically extracted from surrounding text, image caption, page title, and/or image file name. Our work differs from previous in that we focus on identifying, capturing, and combining image context information from a variety of sources. Our goal is to let Usage Context from multiple environments be combined and used together with Capture Context, which collectively will serve as indicators of the semantic content of the image.

4 Combining image context information

As an image (or copies of an image) can be included in several collections or documents, there may be equally many Usage Context descriptions related to the image and each Usage Context description can give an *indication* to what the image is about.

An objective of our work is to develop an automated process, controlled by rules determining interpretation policies, that uses Usage Context data for semantic image interpretation. One approach is to compare Usage Context from different

⁴http://images.google.com

environments (for instance keywords from a collection description and a document abstract) to identify topically overlapping contexts that collectively support a semantic view of an image, or identify topically disparate context that may indicate different semantic views of an image. Rules can be developed to determine the relative importance of context information, how information is compared, and how information is combined to infer image semantics.

To illustrate how a combination of context information can be used, assume an image collection containing images of "Boats", $Coll_{Boats}$, and a document describing a voyage with the ship "Hurtigruten". In this example we represent all images in the document as an image collection, $Coll_{Hurtigruten}$, see figure 1. The document contains different types of images, showing for instance exterior and interior images of "Hurtigruten", sights from different harbors, and activities onboard. However, if an image is included in both $Coll_{Boats}$ and $Coll_{Hurtigruten}$, we may, based on a topic description of the collection, have reason to believe that the image depicts a ship of type "Hurtigruten".

We also assume a third set of images, containing images on the topic "Sights from the Lofoten Islands", $Coll_{Lofoten}$. An image appearing in all three collections is most probably depicting a "Hurtigruten" ship somewhere in Lofoten.

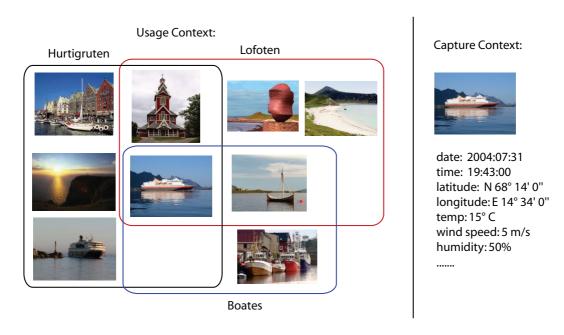


Figure 1: Image contexts

Other types of information, such as the purpose or application of an image, can also be deduced based on Usage Context information. Assume we need an image of a boat to be used as illustration in a document. Among the images in $Coll_{Boats}$, some are suited as illustration, while others are not. If an image is included in both the "Boats" collection and the document about "Hurtigruten", we may have found an image with a relevant topic that is also suited as illustration. Usage Context information can also be useful for determining what the image, most likely, is not about.

Capture context information of different types, for instance date/time, location, temperature and wind speed, provide useful additions to the Usage Context information. If we are searching for an image of "Hurtigruten" at a specific harbor,

location information (such as GPS coordinates) are useful. When searching for "Hurtigruten" on a stormy day, wind speed and humidity may be useful. And a search for "Hurtigruten" in midnight sun might, in addition to Usage context information, use date/time and location as indicators of the possibility of midnight sun.

5 Accumulating context information

Image context information, describing both capture and usage environments, may be automatically collected from different context sources and stored as image metadata using a syntax as found in the CTXT vocabulary presented in Section 6.

At image capture time different types of Capture Context information can be collected, such as the date/time and possibly the GPS coordinates if the image is captured by a device supporting location tagging. Additional Capture Context information can be added later by contacting context sources that hold information describing some aspect of the capture environment, for example from services on the Internet. Available resources include the GeoName⁵ service which takes GPS coordinates as a query and returns geographical names together with a number of other features. EarthTools⁶ will, based on a geographic name, give sunrise and sunset times, local time, time zone, and elevation. The National Climatic Data Center⁷ provides weather observation recordings based on either GPS coordinates or geographical name.

Usage Context information can accumulate over time, since an image may be copied several times and included in different collections or documents. For each new usage environment, Usage Context information can be collected and added to the image metadata.

Context information can be stored either as part of the image file or collected in a context repository. Some Capture Context information, such as location, temperature, wind speed and other types of sensor data, normally apply to a single image and are collected at (or near to) image capture time. This information should be stored in the header of the image file. For context information that applies to a number of images (such as a textual description of a city or building) and for information that accumulates over time (such as Usage Context information), a common context repository could be preferable.

When a context repository is used, all relevant context information, for every copy of an image, should be available in the repository. This implies that the repository is updated with new information for each new Usage Context element that is associated to one of the image copies. However, a context repository including all context information about an image, can only be used if all usage environments can access the repository for adding context descriptions and for use in image retrieval. If this is not possible, Usage Context information and other context information from the repository, can be stored in the header of the image file and distributed with new image copies.

As an example, consider figure 1, and assume that collections $Coll_{Boats}$ and $Coll_{Lofoten}$ reside on site A. The author of the document about "Hurtigruten" is searching for images to illustrate the text. She finds an interesting image, O, in

⁵http://geonames.org

⁶http://www.earthtools.org/

⁷http://www.ncdc.noaa.gov/oa/ncdc.html

collection $Coll_{Boats}$, and downloads a copy of O to site B. However, before O is downloaded, information describing all known context sources, i.e. Capture Context and collections $Coll_{Boats}$ and $Coll_{Lofoten}$, are stored in the header of the image file.

When the image is included in the "Hurtigruten" document, this document represents a new Usage Context for the image, and Usage Context information, such as keywords from the abstract or from surrounding text, is added to the image metadata. If the image is made available for retrieval from site B, an image retrieval system will have access to information about all three Usage Contexts.

The example above illustrates that context information can be inherited from an older copy of the image. This information will then follow every copy of the image. Context information associated with image copies are thus spread in a tree-like fashion. Image copies in the leaves of the tree will contain different sets of context information, which reflect a history of the context information for 'their' copy. The above approach is feasible as long as the header of the image file can be updated with new context information.

Context information can be saved as part of the comment block for JPEG, GIF, and PNG type images. A comment block in a JPEG file can contain arbitrary text, each block has a limited size, but there can be as many blocks as necessary. The Exif⁸ format, which is used by most digital cameras to hold image metadata, can also include context information. In that case, we can use the Comment element in Exif, that can include textual information of any kind.

6 A vocabulary for image context information

We will in this section describe how image context information can be organized and stored as metadata to images. To represent image context data we present a new context vocabulary, CTXT, that is designed for structuring image context information so that reasoning about the different types of context information is possible. A small subset of CTXT has previously been presented in [15].

We base CTXT on RDF⁹, which is a framework for resource description that supports characteristics such as extensibility and interoperability. Extensibility is needed in CTXT since new context information can always be added, and interoperability is useful when context information is used in different environments. CTXT is inspired by both the Dublin Core (DC)¹⁰ and the FOAF¹¹ (Friend of a Friend) vocabularies.

CTXT elements

The CTXT image context vocabulary includes elements for describing both Usage and Capture Contexts. We currently have four classes in CTXT. There is one class for each of the context classes described in section 3; i.e. ctxt:CaptureContext and ctxt:UsageContext. The classes ctxt:SensorData and ctxt:InfoUnits are subclasses of ctxt:CaptureContext, and include respectively information gathered from sensors and information provided by (non-sensor) context sources. For each class we also identify context types. Figure 2 shows the four context classes and gives examples of some useful context types.

⁸http://exif.org/

⁹http://www.w3.org/RDF/

¹⁰http://www.dublincore.com

¹¹http://xmlns.com/foaf/0.1/

Context classes: UsageContext, CaptureContext, SensorData, InfoUnit

Context types:

UsageContext: Collection, Document, Surrounding Text

CaptureContext: Time, Location, Camera SensorData: Location, Weather, Speed

InfoUnit: Time, Area, Object, Person, Event, Action

Figure 2: Context classes and types

Information from each context source will be represented as a context type in the CTXT description of an image. Using the context classes and types from Figure 2, we can register weather, location and speed information from sensor sources, and information (such as textual descriptions) of for instance objects (buildings, monuments), persons and events.

Appendix A lists a subset of the vocabulary elements in CTXT. Some elements identify and describe the context source, and can be used for reasoning about the relevance and usefulness of the source. Other elements can be used for structuring the information provided by the context source. The list of elements related (in particular) to the SensorData and InfoUnit classes, represents only a few examples of useful elements. The wide variety of context information in these classes will result in a longer list of context types and CTXT elements.

Figure 3 illustrates use of CTXT to define usage contexts for the common image in Figure 1. The example is taken from the specifications for the image in the Boats collection, given in Appendix B.

Figure 3: CTXT example

In Figure 3 we have a ctxt:UsageContext description of type <Collection>. The image collection, named "Boats", is public, and we have both an URI to the collection, a textual description of the collection, and keywords describing the collection.

There will be one <ctxt:UsageContext>..... </ctxt:UsageContext> block for each environment where the image is used. In Appendix B we find a ctxt:UsageContext description of type <Document> that includes document title, source identifier, document abstract, keywords and intended audience of the document.

A <ctxt:CaptureContext> </ctxt:CaptureContext> block is used for each type of Capture Context information associated to the image. Application B shows two examples of Capture Context descriptions. The first ctxt:CaptureContext description includes the GPS coordinates at the location of image capture, while the second description includes weather information (temperature, humidity and wind speed) at or nearby the image capture location.

CTXT descriptions can be stored either in a context repository or as part of the comment block of image files, as described in Section 5. Independent of where the context information is stored, the CTXT descriptions should be expandable. As new image contexts are collected, the set of descriptions will be extended.

7 Conclusion

This paper presents 2 classes of context information for images that can be utilized to improve current image retrieval techniques. Capture Context describes the environment in which the image is captured and may include a wide range of data such as capture time, location, and a variety of sensor data. Usage Context describes the potentially multiple environments in which an image is used, such as image collections and documents of various types.

We have presented an RDF based context description vocabulary, called CTXT, which includes constructs for describing both Capture Context and Usage Context information. The CTXT vocabulary is designed for structuring image context information so that reasoning about the different types of context information is possible. The vocabulary is extendable so that new context description constructs can be included as new image contexts are identified.

We argue that combining context information from different sources can enhance our knowledge of the image by representing different and/or complementary views of the image. We believe that this combination of image context information may improve detection of image semantics, and thus may contribute to narrowing the gap between user needs for semantic image retrieval and the capabilities of current image retrieval systems. Future work includes extending CTXT to support more context types, specifying rules and methods for context interpretation, and developing new techniques for efficient use of image context information in image retrieval.

Acknowledgements:

The authors appreciate support from the Norwegian Research Council, NFR, through project nr. 176858, Context-aware Image Management, CAIM, http://caim.uib.no

References

- [1] Rong Zhao and William I. Grosky. Bridging the semantic gap in image retrieval. In Distributed multimedia databases: techniques & applications, pages 14–36, Hershey, PA, USA, 2002. Idea Group Publishing.
- [2] G. Kowalski and M.T. Maybury. *Information Storage and Retrieval Systems Theory and Implementation*. Kluwer Academic Publ, 2nd ed edition, 2000.
- [3] Atsuo Yoshitaka and Tadao Ichikawa. A survey on content-based retrieval for multimedia databases. *IEEE Transactions on Knowledge and Data Engineering*, 11(1):81–93, 1999.

- [4] A. Dey and G. Abowd. Towards a better understanding of context and context-awareness. In Workshop on The What, Who, Where, When, and How of Context-Awareness, Conference on Human Factors in Computer Systems, 2000.
- [5] Adrian Graham, Hector Garcia-Molina, Andreas Paepcke, and Terry Winograd. Time as essence for photo browsing through personal digital libraries. In *JCDL '02: Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries*, pages 326–335, New York, NY, USA, 2002. ACM Press.
- [6] Matthew Cooper, Jonathan Foote, Andreas Girgensohn, and Lynn Wilcox. Temporal event clustering for digital photo collections. *ACM Trans. Multimedia Comput. Commun. Appl.*, 1(3):269–288, 2005.
- [7] Y.-H. Lu and E. J. Delp, III. An overview of problems in image-based location awareness and navigation. In Security, Steganography, and Watermarking of Multimedia Contents VI. Proceedings of the SPIE, Volume 5308, pages 102–109, January 2004.
- [8] Julien Pauty, Paul Couderc, and Michel Banatre. Using context to navigate through a photo collection. In *MobileHCI '05: Proceedings of the 7th international conference on Human computer interaction with mobile devices and services*, pages 145–152, New York, NY, USA, 2005. ACM Press.
- [9] Mor Naaman, Susumu Harada, QianYing Wang, Hector Garcia-Molina, and Andreas Paepcke. Context data in geo-referenced digital photo collections. In MULTIMEDIA '04: Proceedings of the 12th annual ACM international conference on Multimedia, pages 196–203, New York, NY, USA, 2004. ACM Press.
- [10] F. Monaghan and D. OSullivan. Generating useful photo context metadata for the semantic web. In *Proceedings of the 7th International Conference on Mobile Data Management (MDM'06)*, 2006.
- [11] Paul Holleis, Matthias Kranz, Marion Gall, and Albrecht Schmidt. Adding context information to digital photos. In *ICDCS Workshops*, pages 536–542, 2005.
- [12] Kerry Rodden and Kenneth Wood. How do people manage their digital photographs? In CHI 2003 Conference on Human Factors in Computing Systems, April 2003.
- [13] M. Swain, C. Frankel, and V. Athitsos. Webseer: An image search engine for the world wide web, 1997.
- [14] Thijs Westerveld. Image retrieval: Content versus context. In Content-Based Multimedia Information Access, RIAO, 2000.
- [15] R.Karlsen and J. Nordbotten. Capturing diverse usage contexts for web-based images. In ICIW '08: Proceedings of the Third International Conference on Internet and Web Applications and Services. IEEE Press, 2008.

Appendix A: Examples of CTXT vocabulary elements

Element What it means

Elements of general usage:

Sourceld The identification of the context source, for instance an URI.

SourceLocation Identifies the location of the context source.

SourceDescription A textual description of the context source.

Type Specifies the type of the context.

Description A full-text description of context information.

Language The language of the context information.

Keywords Descriptive terms specifically relevant for the context.

Specific elements for Usage Context information:

Title Title of the document or collection where the image occur.

Abstract The abstract of a document.

Owner The owner of the document or collection.

Creator The creator of the document or collection.

Audience Who the collection/document is intended for.

(e.g. Children, Tourists, Students, Scientist)

Category The theme of the collection/document.

(e.g. History, Nature, Architecture, Sports, Travel, Culture)

Availability Availability of the context. It can be public, private or restricted.

Specific elements for InfoUnit:

TimeOfDay Characterization of the time of day.

(e.g. Morning, Noon, Afternoon, Evening, and Night)

Season Characterization of the season (Spring, Summer, Autumn, and Winter).

Month Determines the month of year.

Name of a Location, an Object, a Person.

Specific elements for Sensor data:

Latitude A latitude measurement.

Longitude A longitude measurement.

Altitude The hight above the sea level.

Temperature A temperature measurement.

Humidity The amount of water vapor in the air.

WindSpeed A wind speed measurement.

Format Specifies the format of the sensor measurement

(e.g. Celsius, Fahrenheit, m/sec)

Appendix B: CTXT usage contexts for an image in Figure 1.

```
<?xml version0"1.0" encoding="UTF-8"?>
<rdf:RDF>
   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
   xmlns:dc="http://purl.org/dc/elements/1.1/"
   xmlns:ctxt="http://www.caim.cs.uit.no/ctxt/0.1/"
   <ctxt:UsageContext>
        <ctxt:Type>Collection</ctxt:Type>
        <ctxt:Title>Boats</ctxt:Title>
        <ctxt:SourceId> "http://www.caim.cs.uit.no/Boats" </ctxt:SourceId>
        <ctxt:Availability> public </ctxt:Availability>
        <ctxt:Description>
            "This collection includes images of different types of boats in
           Norway during...."
        </ctxt:Description>
        <ctxt:Keywords> "boat", "Norway",... </ctxt:Keywords>
   </ctxt:UsageContext>
   <ctxt:UsageContext>
        <ctxt:Type>Document</ctxt:Type>
        <ctxt:Title> Hurtigruten - a voyage along the Norwegian coast </ctxt:Title>
        <ctxt:SourceId>
            "http://www.caim.cs.uit.no/Hurtigruten.pdf"
        </ctxt:SourceId>
        <ctxt:Abstract>....</ctxt:Abstract>
        <ctxt:Keywords>....</ctxt:Keywords>
        <ctxt:Audience> tourists </ctxt:Audience>
   </ctxt:UsageContext>
   <ctxt:CaptureContext>
        <ctxt:SensorData>
           <ctxt:Type>Location</ctxt:Type>
           <ctxt:Latitude> N 68 14 0 </ctxt:Latitude>
           <ctxt:Longitude> E 14 34 0 </ctxt:Longitude>
            <ctxt:Altitude> 1 </ctxt:Altitude>
        </ctxt:SensorData >
   </ctxt:CaptureContext>
   <ctxt:CaptureContext>
        <ctxt:SensorData >
           <ctxt:Type>Weather</ctxt:Type>
           <ctxt:SourceDescription> Weather station Model X </ctxt:SourceDescription >
           <ctxt:SourceId> ..... </ctxt:SourceId>
           <ctxt:SourceLocation>
                <ctxt:Longitude> </ctxt:Longitude>
                <ctxt:Latitude> </ctxt:Latitude>
           </ctxt:SourceLocation>
            <ctxt:Temperature>
                <ctxt:Format> Celsius </ctxt:Format>
                <ctxt:Value> 15 </ctxt:Value>
           </ctxt:Temperature>
           <ctxt:Humidity> 50 </ctxt:Humidity>
            <ctxt:WindSpeed> 5 </ctxt:WindSpeed>
        </ctxt:SensorData >
   </ctxt:CaptureContext>
</rdf:RDF>
```