Answering Spatio Temporal Multimedia Queries Using Configurable Ontology

Sunil Kumar Kopparapu, Arun Pande
TCS Innovation Labs - Mumbai, Yantra Park, Thane (West), Maharastra, INDIA
SunilKumar.Kopparapu@TCS.Com

Keywords: spatio-temporal multimedia, question-answering, ontology

Abstract: With the mobile phone becoming ubiquitous it has become easy to find the channel that can assist people in rural areas query for answers from experts in the urban areas. However, the number of experts in any domain are significantly outnumbered by the people seeking information in that domain; this necessitates a question answering (QA) system. In this paper, we discuss a system that is required to answer a query that has a spatial, a temporal and a multimedia component. We concentrate on the agricultural domain due to our previous experience in building a platform (mKRISHI - An Agro Advisory Platform) that enables rural farmers seek expert answers. The focus of the paper is in (a) formulating the multimedia query answering problem and (b) proposing a configurable ontology to assist answering a query.

1 INTRODUCTION

Question answering (QA) systems (Lopez et al., 2007) have been around for a while now and the intent of building these systems is to enable a machine automatically answer a query, generally posed in natural language (Wu et al., 2011). Most of the research work is concentrated on how to derive the intent of the query and then to ensemble and construct an answer from a set of answer paragraphs aided by a domain knowledge base or an ontology.

In recent times, especially with the proliferation of mobile phones it has become easy for folks in the rural regions seek answers from experts in the urban regions on a wide range of topics. From the Indian context, agricultural related information is in high demand where folks in the rural areas expects answers from the experts who are in the urban areas. While a human expert can answer a query without much problem, the relevance of the answer increases with the knowledge of current and past environmental conditions from which the query originated. To enable experts have access to environmental information in addition to the query, we built a platform called mKRISHI, a Mobile Agro Advisory Platform (TCS, 2012), which an expert a view of the environment along with the query so that the expert can answer the query very precisely. The query from the rural folks generally consists of audio, video, photograph in addition to the textual query. In several domains (more so in agriculture) the answers to the same query is different for different spatio-temporal[1] conditions and could be very personalized to the person seeking the answer. In brief, same or similar queries coming from different geographic (spatial) regions could evoke different answers. For example, a query about *Should I apply pesticide now?* coming from a farmer in a region that is expecting mild showers in the next few days need to be addressed differently from the same query coming from a region that has no possibility of rain in the next couple of weeks. Additionally, same query coming from a farmer who own one acre of land and a farmer who own hundred acres of land needs to answered differently. In such a scenario, very often the answers that an experienced expert would provide would be based on several parameters (spatio-temporal) including specific personal details of the person seeking the answer.

Like in any scenario, the number of human experts are considerably outnumbered by the people seeking information, this necessitates the need for an automatic question answering system. Question Answering (QA) systems are being increasingly finding use in all walks of life. Domain specific QA systems address the task of automatically answering a question posed in natural language and generally deal only

[1]We refer to the latitude and longitude as the spatial information and temporal refers to the time of the query in terms of time of the year etc.

with questions in a specific domain (for example, a QA system for medicine is not expected to answer a query related to automotive maintenance). In many cases domain specific QA systems exploit domain-specific knowledge frequently formalized in ontologies (WikipediaQA, 2012) to function. An ontology formally represents the domain knowledge as a set of concepts, and the relationships between them. These concepts and the relationship between them can be exploited to not only describe the domain but also enable extraction of relationship between different entities within that domain (WikipediaOnt, 2012; Xie et al., 2008). Generally ontologies, especially in the area of agriculture, may be unable to provide personalized level of information. More recently (Bansal and Malik, 2011) provide a framework for crop production life cycle which not only provides relevant but contextual relevant as well as scientifically correct information. However, in an ideal situation a QA system (assisted by a suitable ontology) assumes that the query posed is perfect in the sense of completeness of the information in the query. In reality, the queries are often incomplete and fuzzy. In the absence of an ontology being able to assist address such fuzzy queries would end up making the QA system unable to address the query and/or prone to erroneous answers.

In this paper, we first formulate a multimedia query answering problem and then propose a reconfigurable ontology that assists in answering the multimedia query. The rest of the paper is organized as follow, in Section 2 we formulate the problem and provide a solution approach in Section 3. We give some experimental results in Section 4 and conclude in Section 5.

2 PROBLEM FORMULATION

Given a multimedia query consisting of text, image, video, speech along with spatial and temporal information of the past and future (predicted). We try to address the problem of how to answer the query precisely?

For the sake of clarity, we assume the query to be related to agricultural domain with the understanding that this is in no way restricted to the agricultural domain specifically. Let q_t be the query posed at time t and let q_t in addition to the query text consist of

a Spoken data, S_t, at time t (spoken query in natural language),

b image data, I_t at time t (image or the photograph of the crop related to the query),

c video data, V_t at time t (video of the crop or a farming procedure related to the query),

d spatial location L (the location from where the query originated),

e farmer data P (details of the farmer in terms of land holding),

f farm data F_k for $k \in [t_0, t]$ for location L (details of the farm like agricultural operations done, pesticides and fertilizers used to date),

g weather data W_m for $m \in [t_0, t] \cup]t, t+\delta]$ for location L (like temperature, rainfall obtained from sensors, both past and predicted),

Here δ is the time interval for which prediction data is available and t_0 is some initial finite start time since when information is available. So the spatio-temporal multimedia query q_t can be represented as

$$q_t = (S_t, I_t, V_t, P, L, \{F_k\}_{k=t_0}^{t}, \{W_m\}_{m=t_0}^{t+\delta}), \quad (1)$$

now the problem is that of finding a QA system $\mathcal{F}$ which can use the knowledge $\mathcal{K}$ to identify an answer $\mathcal{A}_t$, namely,

$$\mathcal{A}_t = \mathcal{F}(q_t / \mathcal{K}) \quad (2)$$

For the sake of making this problem tractable, we assume that it is indeed possible to represent the query in multimedia in the form of text. For example,

$$S_t \xrightarrow{\text{speech recognition}} \mathcal{T}_{S_t}$$

where the spoken query S_t can be converted into text (say into $\mathcal{T}_{S_t}$) using an automatic speech recognition engine (Imran and Kopparapu, 2011), and

$$I_t \xrightarrow{\text{image interpretation}} \mathcal{T}_{I_t}$$

where the image I_t can be converted into text (say $\mathcal{T}_{I_t}$) using image processing followed by pattern matching and image analysis, similarly all the spatio-temporal data can be represented in the form of text. Namely,

$$\mathcal{T}_{q_t} = \Big(\mathcal{T}_{S_t} \cup \mathcal{T}_{I_t} \cup \mathcal{T}_{V_t} \cup \mathcal{T}_{\{F_k\}_{k=t_0}^{t}} \cup \mathcal{T}_P \cup \mathcal{T}_L \cup \mathcal{T}_{\{W_m\}_{m=t_0}^{t+\delta}} \Big) \quad (3)$$

Now the problem (2) is one of given a set of text terms, $\mathcal{T}_{q_t}$, derived from different forms of multimedia data, find an answer $\mathcal{T}_{\mathcal{A}_t} (= \mathcal{A}_t)$ given the domain knowledge $\mathcal{T}_{\mathcal{K}} (= \mathcal{K})$, namely,

$$\mathcal{A}_t = \mathcal{F}(\mathcal{T}_{q_t} / \mathcal{K}) \quad (4)$$

2.1 Multimedia Query ($\mathcal{T}_q$)

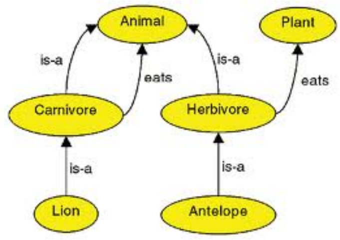
A typical query is shown in Table 1. The first three data items (S_t, I_t, V_t) are sent by the person posing the query and the rest are obtained automatically, for example W_t is obtained from the sensors on the field (TCS, 2012). Note that $\mathcal{T}_{q_t}$ is the union of all the derived text in Table 1. Namely, $\mathcal{T}_{q_t} =$

Data	Sample	Derived Text ($\mathcal{T}_*$)
S_t		I have been having white spots on my grapes what could that be?
I_t		Black Grapes have mild powdery spots visible
V_t	None	–
L	Automatic	12 km north of Nasik
P	Automatic	Ramlal is a knowledgeable farmer, but poor
t	Query Time	Harvesting season in Nasik
F_t	Automatic	Used new seed variety; sprayed Potassium last week
W_t	Automatic	Warmer than usual for this time of the year
W_{t-}	Past Data	Has been cold in the last two days
W_{t+}	Predicted Data	Likely to rain with probability 0.9 in 24 hours

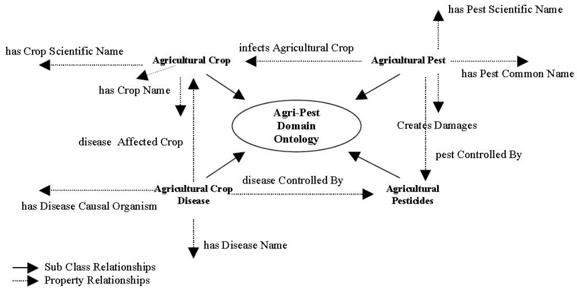
Table 1: Typical Multimedia Query

$$\left\{\begin{array}{l}\text{I have been having white spots on my ... that be?}\\ \text{Black Grapes have mild powdery spots visible}\\ \text{12 km north of Nasik}\\ \text{Ramlal is a knowledgeable farmer, but poor}\\ \text{Harvesting season in Nasik}\\ \text{Used new seed variety; sprayed Pottas .... week}\\ \text{Warmer than usual for this time of the year,}\\ \text{Has been extremely cold in the last two days,}\\ \text{Likely to rain with probability 0.9 in 24 hours}\end{array}\right\}$$

Note that in reality the process of converting a multimedia input (especially, S_t, I_t, V_t) is not very accurate. For example conversion of $S \rightarrow \mathcal{T}_S$ requires a speech recognition of a naturally spoken query by the user. The automatic speech recognition (ASR) accuracy is very poor for resource deficient languages especially when the system has to work in a speaker independent mode. The same holds good for converting $I \rightarrow \mathcal{T}_I$ or $V \rightarrow \mathcal{T}_V$ which would require significant amount of image and video analysis respectively. These areas, namely, speech recognition, image interpretation (Deruyver et al., 2009) and video interpretation (Selman et al., 2011) are being actively addressed by researcher around the globe. For the purpose of our discussion we will assume that the process of converting multimedia data into the corresponding text is error free.



(a) Basic ontology representation.



(b) A agri-pest domain ontology (Urs and Angrosh, 2007) .

Figure 1: Domain ontology representation.

2.2 Knowledge ($\mathcal{T}_\mathcal{K}$)

The domain knowledge, $\mathcal{T}_\mathcal{K}$, is essentially an ontology which by definition is the formal representation of the knowledge in terms of a set of concepts within a domain, and the inter relationships between them (Wikipedia, 2012). A sample $\mathcal{K}$ is given in Fig. 1. While 1(a) shows a sample generic ontology, Fig. 1(b) shows a specific agri-pest domain ontology. In a very loose sense, $\mathcal{K}$ can be looked upon as a graph, where the nodes are the concepts related to the domain (example, `harvest`, $\cdots$) which are related to the entities (for example, `paddy`, $\cdots$) and the edge connect the concepts and the entities. However, the edges can be associated with a strength, in a weak sense, a probability measure that captures the degree of association between the entity and the concept. As we will show in the next section, these edge weights in an ontology are dependent on the spatio-temporal information in a multimedia query.

2.3 Query Answering ($\mathcal{T}_{\mathcal{A}}$)

The problem is now that of solving (4), namely finding, a $\mathcal{F}$, such that a $\mathcal{T}_{\mathcal{A}}$ can be identified for the query $\mathcal{T}_q$ using $\mathcal{K}$ or alternatively, finding a set of $\mathcal{T}_{\mathcal{A}}$ that are ranked in a way to best address $\mathcal{T}_q$. Note that $\mathcal{F}$ could be, typically a Question Answering System. Typically, one would expect $\mathcal{T}_{\mathcal{A}}$ as

> *Spray insecticide at* 1 *ml per* 1 *litre of water after* 48 *hours.*

in response to the multimedia query q_t.

3 SOLUTION APPROACH

This problem formulation is in the form of a typical Question Answering (QA) system (Kopparapu et al., 2007). A natural language text query is the input to the QA system and the system is to identify a relevant set of possible answers corresponding to the query from a set of answers that is already known to the system.

However, in this paper, we adopt a slightly different approach to address this problem. We visualize the answer to lie in the structure of $\mathcal{K}$ itself; however unlike the typical ontology, we use a reconfigurable $\mathcal{K}$ (Pande and Kopparapu, 2011) which configures itself based on the query there by leading to the correct answer selection.

3.1 Reconfigurable $\mathcal{K}$

The central idea is to construct $\mathcal{K}$ dynamically, the dynamism being triggered by the nature of the query, the time of the query, the origin of the query, etc. The idea being that a dynamic $\mathcal{K}$ captures the context more precisely and hence is able to assist the QA system point to an appropriate answer quickly and reliably. In essence $\mathcal{K}$ captures the environmental conditions that exists at the farmer's location and this enables answering the spatio-temporal multimedia query. Essentially $\mathcal{K}$ dynamically adjusts itself so that it is most relevant to the posed query. This dynamism assists the QA system to perform with higher precision and hence imitating an expert to identify the most *relevant* answer. As mentioned earlier, $\mathcal{K}$ can be looked upon as a graph, where the nodes are the concepts related to the domain (example, `harvest`, ...)

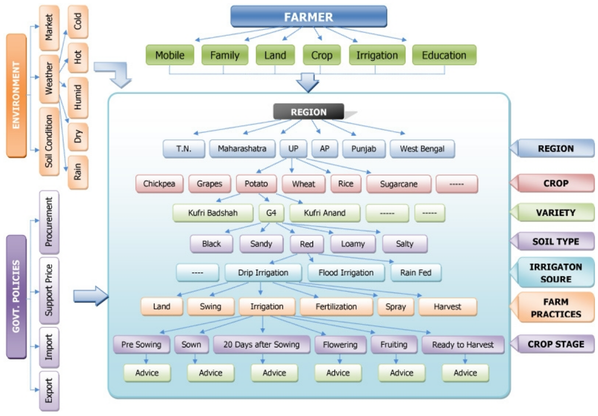


Figure 2: A sample $\mathcal{K}$

which are related to the entities (example, `paddy`, ...) and the edge connecting the concepts and the entities are associated with a strength, in a weak sense, a probability measure. This strength of the edge connecting the concept and the entity is dependent, for example, on the time of the year (season) when the query was asked. For example all the entities connecting to the concept `harvest` would be fully connected (probability 1) in the actual harvesting season (say if the query came in September) while all connections to the concept `harvest` would be broken (probability 0) if the same query were to come in say December (non-harvesting season). Similarly, the connectivity between the concept `harvest` and the entity `paddy` could be 0 for a query coming from a non paddy growing region (say Punjab) and 1 when the query comes from a paddy growing region (say West Bengal). In this particular example, the strength of the connectivity between the concept `harvest` and the entity `paddy` would be the product of all these conditional probabilities, in this example, season and region. Namely the strength of the edge between `harvest` and `paddy` would be $P(\text{season}) \times P(\text{region})$. As another example, the association between `spray` and `insecticide` would be based on the probability that it would rain in the next couple of days. A higher probability of rain would decrease the association while a prediction of no rain would strengthen the association. A typical $\mathcal{K}$ for agricultural practices is shown in Fig. 2.

4 EXPERIMENTAL RESULTS

We handcrafted $\mathcal{K}$ as shown in Fig. 2 with the help of agricultural experts. We extracted about 100 real queries that the farmers has asked and the actual answers given by the human expert on our platform (TCS, 2012). The queries consisted of multimedia information which we converted into text manually by listening to the audio query and by looking at the images associated with the query. In all the cases the response of the expert was a text message. For each of the query, we also extracted the environmental condition information and also the time of the query etc from logs. In almost 90 % of the cases we were able to derive the response from $\mathcal{K}$. For example, an actual response of the expert say "*Spray insecticide at* 1 *ml per* 1 *litre of water after* 48 *hours.*" gave us "*Do nothing*" because the concept rain removing the connection between insecticide and activity of spraying. The experimental results are encouraging and we are in the process of studying further the reconfigurability of $\mathcal{K}$ so as to enable multimedia question answering.

5 CONCLUSIONS

Multimedia queries are common in a variety of domain more so in agriculture domain because of the need for the expert to ascertain several environmental conditions to be able to give a reliable answer. Very often the answers are different for the same query just because the query came from a certain region or at a certain time of the year. In this paper, we first formulated the problem as a multimedia query answering system and then presented a self configuring $\mathcal{K}$ based on the query itself, this is the main contribution of the paper. Preliminary results, shows that this approach is usable in addressing multimedia queries, provided that the multimedia data is converted into text data without any error. We plan to study the problems that can crop up when errors are introduced when multimedia query is converted into text.

REFERENCES

Bansal, N. and Malik, S. (2011). A framework for agriculture ontology development in semantic web. In *International Conference on Communication Systems and Network Technologies*, pages 283 – 286.

Deruyver, A., Hodé, Y., and Brun, L. (2009). Image interpretation with a conceptual graph: Labeling oversegmented images and detection of unexpected objects. *Artif. Intell.*, 173(14):1245–1265.

Imran, A. and Kopparapu, S. (2011). Building a natural language hindi speech interface to access market information. In *Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), 2011 Third National Conference on*, pages 58 –61.

Kopparapu, S. K., Srivastava, A., and Rao, P. V. S. (2007). Minimal parsing key concept based question answering system. In Jacko, J. A., editor, *HCI (3)*, volume 4552 of *Lecture Notes in Computer Science*, pages 104–113. Springer.

Lopez, V., Uren, V. S., Motta, E., and Pasin, M. (2007). Aqualog: An ontology-driven question answering system for organizational semantic intranets. *J. Web Sem.*, 5(2):72–105.

Pande, A. and Kopparapu, S. K. (2011). A self configuring knowledge base representation, indian patent number 3294/mum/2011.

Selman, J., Amer, M. R., Fern, A., and Todorovic, S. (2011). Pel-cnf: Probabilistic event logic conjunctive normal form for video interpretation. In *ICCV Workshops*, pages 680–687. IEEE.

TCS (2012). http://www.tcs.com/offerings/technology-products/mkrishi/pages/default.aspx.

Urs, S. R. and Angrosh, M. A. (2007). Ontology-based knowledge organization systems in digital libraries: A comparison of experiments in owl and kaon ontologies. `www.vidyanidhi.org.in/shaliniurs_files/icdl.pdf`.

Wikipedia (2012). http://en.wikipedia.org/wiki/ontology.

WikipediaOnt (2012). http://en.wikipedia.org/wiki/ ontology_computer_science.

WikipediaQA (2012). http://en.wikipedia.org/wiki/ question_answering.

Wu, G.-L., Su, Y.-C., Chiu, T.-H., Hsieh, L.-C., and Hsu, W. H. (2011). Scalable mobile video question-answering system with locally aggregated descriptors and random projection. In *Proceedings of the 19th ACM international conference on Multimedia*, MM '11, pages 647–650, New York, NY, USA. ACM.

Xie, N., Wang, and Yang, Y. (2008). Ontology-based agricultural knowledge acquisition and application. I 258:349– 357.