Summarization of Broadcast News Video through Link Analysis of Named Entities

Norman Papernick and Alexander G. Hauptmann

Dept. of Computer Science
Carnegie Mellon University,
Pittsburgh, PA 15213
{norm+,alex+}@cs.cmu.edu

Abstract

This paper describes the use of connections between named entities for summarization of broadcast news. We first extract named entities from a transcript of a news story, and find related entities nearby. In the context of a query, a link graph of relevant entities is rendered in an interactive display, allowing the user to manipulate, browse and examine the components, including the ability to play back video clips that mention with interesting relationships. An evaluation of the approach shows that completely automatic summaries from a year of broadcast news can reflect almost 50% of the entities in a manually created reference summary found on the web. Locations are the most accurate aspect of summarization.

Introduction

Informedia research at Carnegie Mellon University has demonstrated the application of speech recognition, image processing, and natural language processing for video analysis and retrieval using terabytes of news spanning years of daily broadcasts [1]. While one aspect of Informedia work emphasizes *retrieval* of relevant broadcast news stories from automatically processed, archived collections [2], in this paper we focus on *summarization* of result sets of documents though link analysis of named entities. While simple objects and named entities that are extracted from the broadcast can be seen as the building blocks of contextual knowledge, this knowledge is insufficient without some understanding of the relations between the entities. Our intent is to create illustrative and informative summaries that capture the core information content and effectively display this information to a user. The user can then drill down to specific passages inside news stories to obtain the original relevant video sequences. Link analysis has previously been applied to extraction and visualization of text news in a number of publications [15]. Earlier work has been done on link analysis of entities in free form documents [13], but the scope had been limited to detecting entities and events. Baldwin and Bagga [13] also limit the context to a single sentence, while our system spans sentence boundaries. Other systems are directed at finding common entities in highly structured data sets. [14]

Extracting Named Entities and their Links

In this section we explain how we extract named entities, unify different surface forms of the same entity into a pseudo-canonical form, and find link information. Our source data is digitally encoded video, which has been processed with automatic speech recognition [16] to create an errorful transcript and video OCR [17] to extract text readable on the screen.

Finding Named Entities

Named entity extraction of people, organization, and location from broadcast news speech transcripts has been done by MITRE via Alembic [8], and BBN with Nymble [9, 10]. Similarly, our processing starts with training data where all words are tagged as people, organizations, locations, or something else. We use a statistical language modeling toolkit [11] to build a tri-gram language model from this training data, which alternates between a named-entity tag and a word, i.e. **–none-** *a* **–none-** *probation* **–none-** *office* **–none-** *in* **–location-** *New York* **–time-** *tomorrow* **–none-** *will* **–none-** *set* **–none-** *restrictions* **–none-** *Martha* **–none-** *Stewart* **–none-** *must* **–none-** *live* **–none-** *with*. To label named entities in new text, we first build a lattice from the text where each text word can be preceded by any of the named-entity tags. A Viterbi algorithm then finds the best path through the named-entity options and the text words, just like speech recognition hypothesis decoding. Thus we can identify locations, organizations, names, dates, and numbers that occur in the transcription of the data streams. All of the extracted named entities and phrases can be time-stamped and synchronized with hyperlinks back to the original multimedia source, making it possible to verify the original content.[7]

Deriving Canonical Forms of Named Entities

Once we have identified the entities, we find that many different surface forms refer to the same entity, especially for names. Thus George Bush, George W. Bush and Bush in the same story refer to the same person, not three different people. To effectively create summaries, we have to unify these different forms into one canonical name.
The following core algorithm is used to map the different forms into a common representation.

1) For each found named entity inside a news story, if an explicit representation for this entity exists in an external knowledge base (e.g., President Bush → George Bush), use that canonical form.
2) If other entities of the same type in the current *story* subsume the current entity as a subset, replace this entity name with the more complete one (e.g., Bin Laden → Osama Bin Laden, if both occur in one story).
3) If other entities of the same type in the current *complete news program* subsume the current entity, replace this entity name with the more complete one (e.g., Bin Laden → Osama Bin Laden, if both occur in one news program).
4) If the current form of the named entity has been referred to more than once in the history of our process, insert it into the database as a "canonical entity".

This process, while conservative, helps merge some of the similar surface forms of entities, and allows input from an external knowledge source. Because the process is automatic and conservative, there are a number of named entities that occur in multiple "canonical" surface forms in the database.

Because the process is derived from speech recognized text or video OCR of fragments of overlaid text, a grammatical parse is impossible, nor can we depend on spelling, capitalization or punctuation. We must accept that there will be substantial numbers of misidentified entities, as well as ones that cannot be automatically merged into canonical form.

Connecting Named Entities

In this subsection, we describe our approach to understanding the relational context of named textual entities, which can be viewed as a form of link analysis.

Informally, two entities are said to be connected if they are mentioned in close proximity. Formally, we can define *e* as a unique instantiation of some named entity, *S* as a single sentence that is represented as a collection of named entities, and *P* as a paragraph which is an ordered collection of sentences. Given that, we can compute the distance between any two named entities as the differences between the indexes of the sentences that contain the entities.

$$S = \{e_o \ldots e_n\}$$

$$P = \{S_o \ldots S_m \ni S_i \cap S_j = \phi\}$$

$$sen(e) = j \text{ such that } e \in P_j$$

$$dist(e_i, e_j) = 1 + \left| sen(e_i) - (e_j) \right|$$

$$weight(e_i, e_j) = \frac{1}{(dist(e_i, e_j))^2}$$

After the distance has been computed, the weight of the connection is computed as the square of the inverse of the distance. For this evaluation we index all pairs that have a distance of 3 or less. The set of all connections can be seen as a graph where the entities are nodes and the weights represent the edges.

Interface Features

The interface around the extracted named entities and their link information affords a variety of presentations and manipulations:

Initial Query. The *Named Entity Connection Graph* viewer is driven by the Informedia Digital Video Library search interface. The search can be text driven (such as "Martha Stewart"), image driven (Query by Image Content), or map driven (e.g., countries in North Africa). Once a set of documents is selected, the *Named Entity Connection Search Engine* returns a list of all the connections within the set. The set can be dynamically filtered using many different techniques. Figure 1 shows the date slider in effect, reducing the number of active documents to only those inside the selected range.

Ranked Ordered List of Entities. The middle right hand portion of Figure 1 contains a scrolling list of named entities. The list is sorted in descending order based on the sum of the weights for all the connections for a given entity. Each entry is color coded based on a gradient from red to blue to represent the relative weight in relation to the other entries. Red represents the largest sum weight, and blue represent the lowest. Clicking on an entry will focus the graph on that named entity.

$$weight(e_i) = \sum weight(e_i, e_j)$$

Summary Graph View. The *Summary Graph View* as shown in the bottom half of Figure 1 shows an overview of how the named entities relate within a search result set. The program Dot [12] is used to generate the graphs. The interface uses colors, shapes, and rollovers to display the relationships. Rectangles represent people, hexagons represent organizations, and ovals represent locations. A

yellow fill indicates the *Entity of Focus*, in this example "Martha Stewart". Cornflower blue filled shapes are those entities which have a direct connection to the Entity of Focus, for this example a maximum distance of 3 is the cutoff. Not shown in Figure 1 are shapes with white fill. These represent the second tier of connections, entities that are strongly connected to the blue entities. The second tier entities are those which would be found in a breadth-first search with depth two, starting at the Entity of Focus. These Entities are related to the Entity of Focus, but not mentioned in close proximity. A directly connected Entity is displayed on in the graph if the Entity is connected inside the subgraph by two or more segments. This removes the noise due to broadcast news "teasers" that contain a summary of the hour's news stories as well false connections due to unrelated stories being mentioned in close proximity.

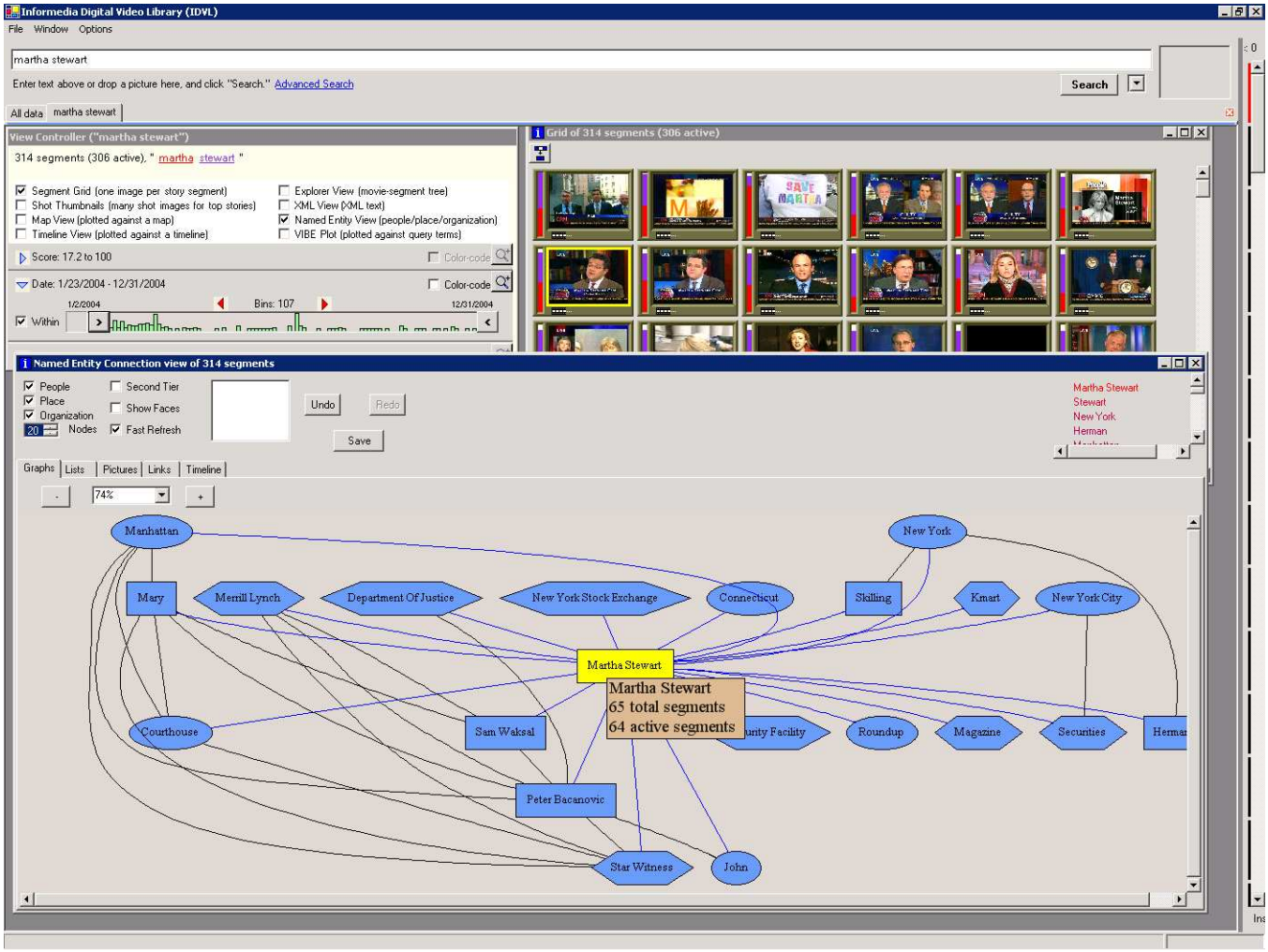

Figure 1. Summary Graph View for the initial query "Martha Stewart."

List View. The List View represents the same connection relations as the Summary Graph View, but in a more compact format as seen in Figure 2. Both the Summary Graph View and the List View share the same color fill idiom. The List View breaks each entity type into a separate column instead of representing type with shape. The actual links between entities are shown when the user moves the cursor over the entity. The selected entity is underlined while the set of connected entities are changed to a bold face.

Text Link View. The Text Link View is activated when the user selects the "play" method from the Graph or List view. The view displays the collection of sentences that are active in the graph with respect to the selected Entity, as seen in Figure 3.

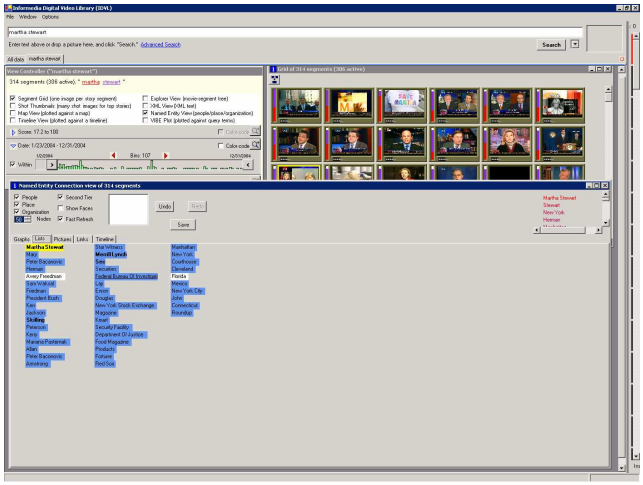

Figure 2. List View

Skim View. A video skim is a temporal multimedia abstraction that incorporates both video and audio information from a longer source. A video skim is played rather than viewed statically, e.g., a ten-minute video segment might be condensed to a one-minute skim [3, 4]. Skims can better summarize audio content and temporal flow than storyboards of single thumbnail images, one thumbnail per shot. For example, important narrative from a single speaker could be captured in a skim, but the storyboard would only show an image of the speaker. As another case, an airplane looping in the sky might be collapsed to a single shot and single image storyboard (especially if the color characteristics remain constant), but a skim could show the interesting motion of the airplane. [5] The Skim View differs from the traditional skims by representing a multi-movie video summary of the connections related to a single named entity instead of focusing on a single movie broadcast. The Skim View uses the precomputed media offsets for the start and end of each sentence to select where in the media the play list should start and stop.

Figure 3. Text Summary View

Afforded Interface Manipulations

The Graph View display allows the user to manipulate the data in many ways. Single left clicking on a shape changes the selected entity to be the new entity of focus. The graph is recomputed with the new entity as the center and displayed to the user. A single right click on a shape brings up a context menu. The entity selected will be referred to as the local focus. The user may *delete* the local focus from the graph. All connections that are dependent on only the local focus are hidden from the display. The user can undelete the node at any time to bring the connections back to view. For instances when the *Canonical Named Entity* process misses due to cases such as "Martha Stewart" and "Stewart" being detected as different entities, the user can *merge* the local focus with any other entity. The target entity's connections are absorbed into that of the local focus. The user can activate the Skim View in two ways from the local focus through *play single* or *play all*. Play single will select all the sentences represented in the current graph that are connected to the local focus and a selected entity. Play all selects all the sentences that are connected to the local focus in the current graph. All right click options are available in both the Graph and the List Views.

In addition to the right click menu, the user is given a set of buttons to interact with the display. The user can zoom into or out from the graph. By default, the system will resize the screen to the smallest size required to display the entire graph without scrolling. Zooming allows the user to inspect a complex graph closely after seeing an overview. The user is given three filters, one for each entity type of people, organizations, and places. Toggling each filter will add or subtract all entities of that type from the connection graph. There is also a toggle for the second tier of connections. When activated, the graph will attempted to include indirectly connected entities after adding all possible directly connected entities. The user may also change the cutoff for the maximum number of entities on the screen. By default, the number is 15.

The user can actively explore the data by highlighting different objects on the screen. When the Summary Graph is active, highlighting a thumbnail of an image as in the upper right hand sections of the screenshots will cause the links that represent the same documents to highlight in the graph. Likewise, highlighting one of the color bars that represent sentences in the Skim View as is shown in Figure 4 will cause the related links to highlight. We refer to this act as "Brushing" and is used in other parts of the Informedia interface.

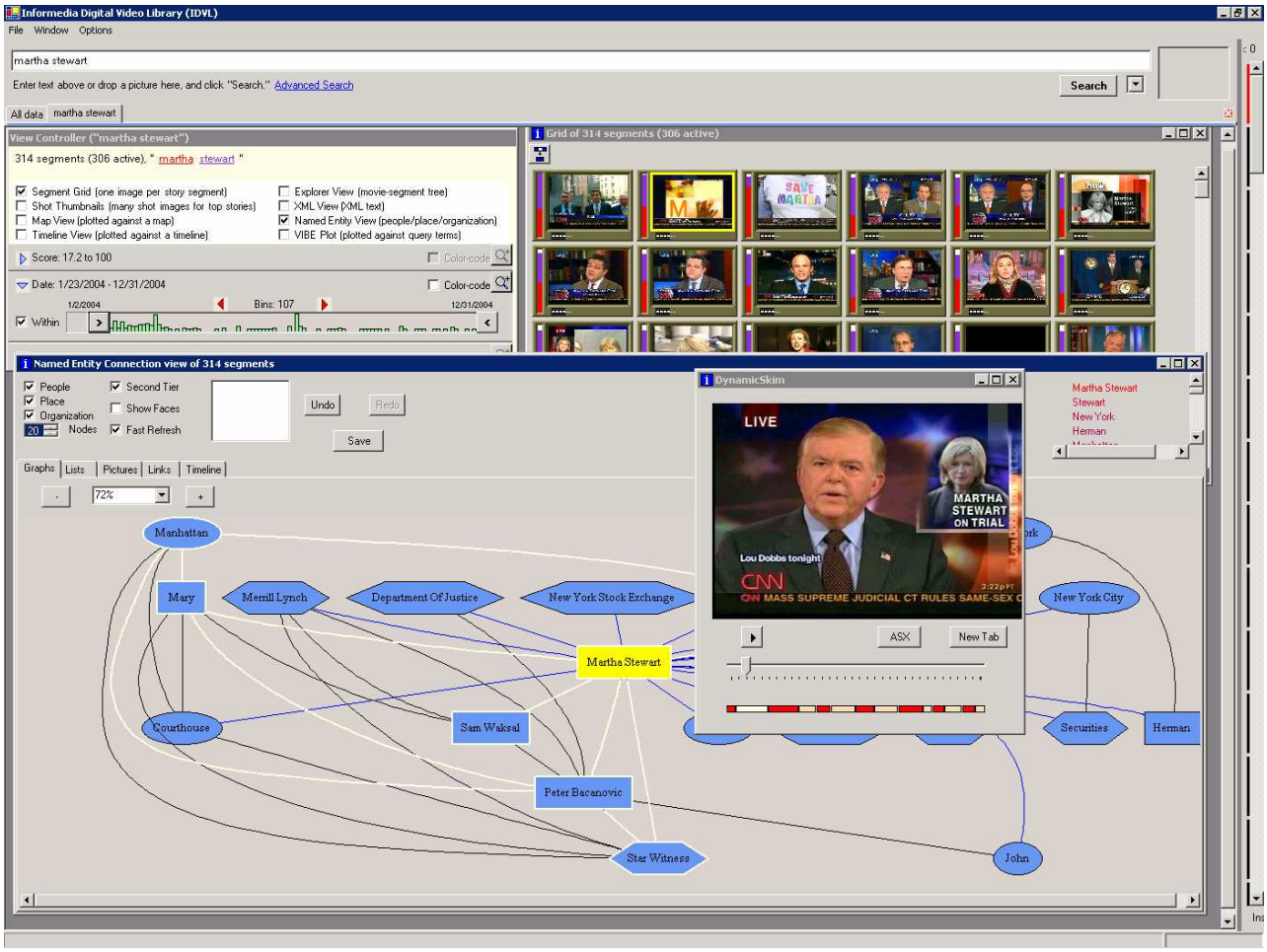

Figure 4. Skim View with highlighting

Evaluation

The Named Entity Connection graph creates a summary of the important related entities for a given entity. To evaluate this claim, we used infoplease.com's list of "People in the News" for 2004 [6]. Each of the 93 entries contains a short description of why that person was in the news that year. An example is the entry for Martha Stewart:

> ***Martha Stewart,*** — diva of domesticity, was sentenced to five months in prison in July after being found guilty on four counts of obstruction of justice and lying to federal investigators. The judge stayed her sentence pending appeal, but ***Stewart*** opted to begin serving her sentence in October. She was also fined $30,000. The charges stem from her December 2001 sale of 3,928 shares of the biotech stock **ImClone**. She made the trade the day before the **FDA** announced it had declined to review **ImClone's** new cancer drug—news that sent shares tumbling.

In this example, there are three entities found (Martha Stewart, FDA and ImClone) with four instantiations (Martha Stewart, Stewart, FDA and ImClone). The test corpus used for the evaluation contained 622 hours of CNN broadcast news from 2004 segmented into approximately 30,000 story documents by Informedia processing. (Table 1)

Table 1

Video Collection in Test Suite	Hours
CNN Lou Dobbs Moneyline	260
CNN NewsNight with Aaron Brown	260
CNN Saturday	51
CNN Sunday	51

For the evaluation, a text search for the given entity was done for each of the 93 names listed by Infoplease for *People of 2004*. A text search could return anywhere from 0 to a configured maximum of 1000 documents. This set of documents was then used to generate the graphs used for evaluation. Only 59 of those contained at least one connection between entities which could be displayed as a graph. The generated graphs were set to a maximum cutoff of 50 entities on the screen. Each graph was generated twice, once with the Second Tier option off and once with the Second Tier option on. The baseline, or First Tier, option limits the graph to only those names that are directly connected to the entity of focus. The Tier 2, or Second Tier, allows for a greater context by considering the entities that are directly connected to each of the linked entities. It contains entities that would be found in a breadth-first search with depth two, starting at the Entity of Focus. The entities were evaluated for accuracy and error rate using the Infoplease as truth. Two different evaluation techniques were used:

(a) Considering each multi word entity as a single **phrase** requiring an exact match and

(b) Treating the list of Named Entities as a **bag of words** to allow for a flexible match.

Given the truth set T and retrieved set R for some Named Entity, the accuracy and error are computed as follows

$$accuracy(T,R) = \frac{|T \cap R|}{|T|}$$

$$error(T,R) = 1 - \frac{\left(|T| - |T \cap R|\right) + \left(|R| - |T \cap R|\right)}{|T| + |R|}$$

The accuracy represents the total overlap between the Infoplease truth set and a given named entity graph. This is directly proportional to recall. The error term is directly related to 1 minus the precision.

Results. Figure 5 shows a leveling off of accuracy of about 37% at 18 entities for exact word evaluation. The additional context of the Second Tier levels off at around 26 entities with an accuracy of 42%.

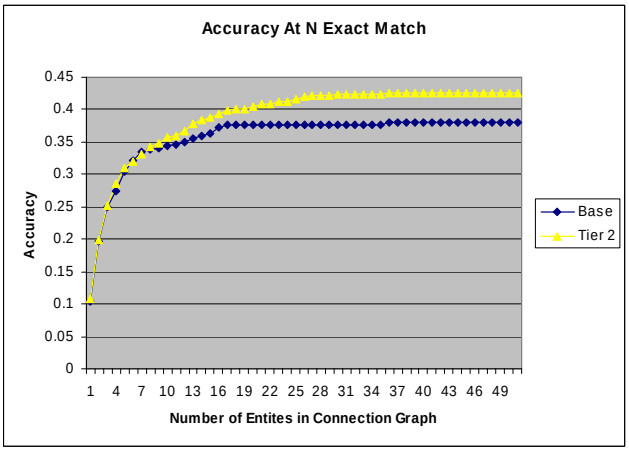

Figure 5. Accuracy for Exact Match direct connections and Second Tier.

The error rate (Figure 6) of the Second Tier is lower at the smaller cutoffs, but as more entities are allowed to be added the improvements become marginal. In terms of error, the Second Tier diverges at about 13 and continues to rise.

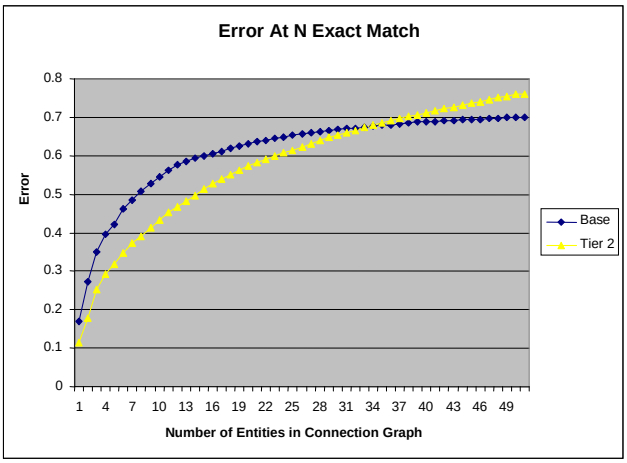

Figure 6. Error for Exact Match direct connections and Second Tier.

In the bag of words evaluation (Figure 7) the accuracy of the direct-only connections levels off at 18 entities with 46% word coverage. The Second Tier continues to rise in accuracy at up to 46 entities on the screen for an average accuracy of about 53%.

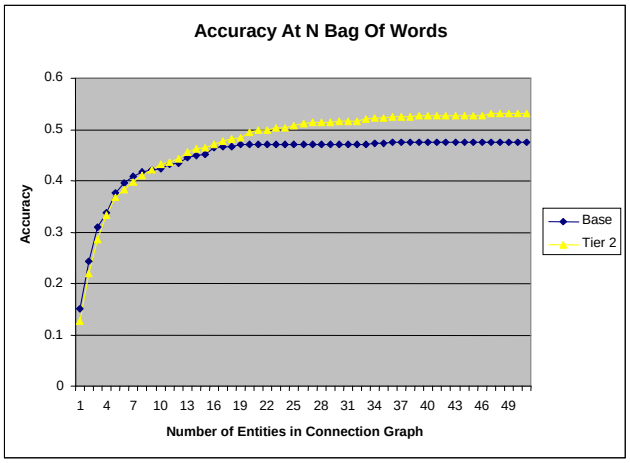

Figure 7. Accuracy for Bag of Words direct connections and Second Tier.

Figure 11 shows a major difference between the human generated summaries and the automatically generated text. The human generated summaries are shorter, containing on average six named entities. The named entity connection based summaries contains more than twice that of the baseline for all cases.

The error for the Bag of Words evaluation stays in lock step for direct-only connections at second tier up to about 12 entities in the list. The second tier error rate diverges as the absolute number of entities being considered is increased (Figure 9).

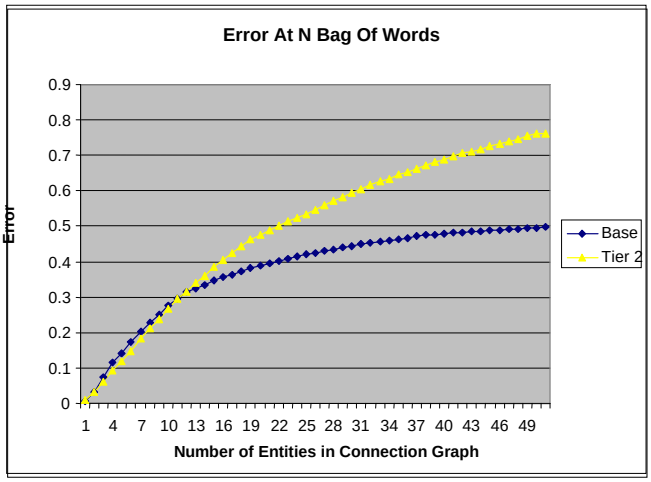

Figure 8. Error for Bag of Words direct connections and Second Tier

The shape of the accuracy and error graphs are explained by the frequency graph in Figure 9. For any given entity, there is a higher chance that the Second Tier graph contains more elements than the direct graph.

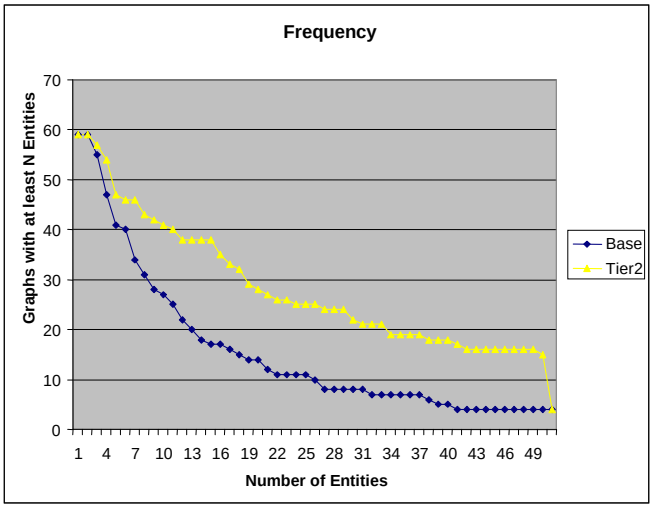

Figure 9. The frequency of number of entities in the automatically generated graphs.

Figure 10 shows the average ratios of each entity type in each evaluation. The named entity connection summaries are evenly distributed between the three types, whereas the truth tends to have more people and fewer locations. Because there were fewer number of organizations in the truth set, organizations are marginally harder to match for most of the situations (Figure 12) Locations had a higher accuracy for the exact match because there are fewer variations in the presentation of location names in media. Names match well in the bag of words evaluation because the truth will sometimes contain an entity listed twice, once with the full name and once with only the last name.

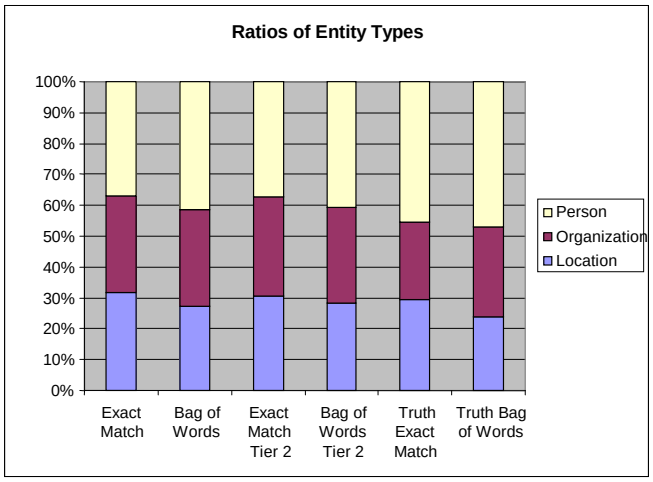

Figure 10. Ratios of entity types for each evaluation category.

The human generated summaries, on average, are much shorter than the automatically generated ones, as seen in Figure 11.

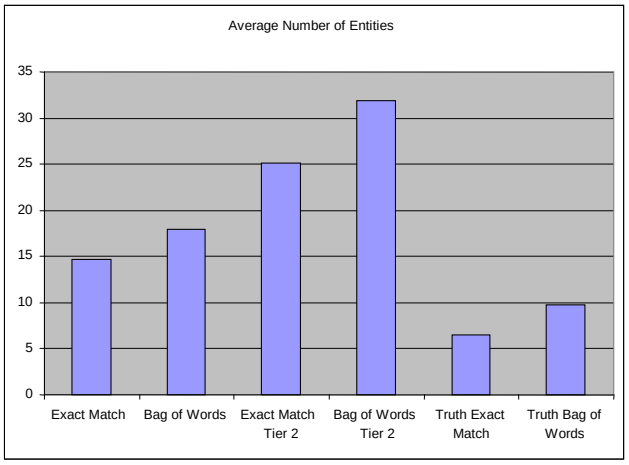

Figure 11. Average number of entities for each evaluation category.

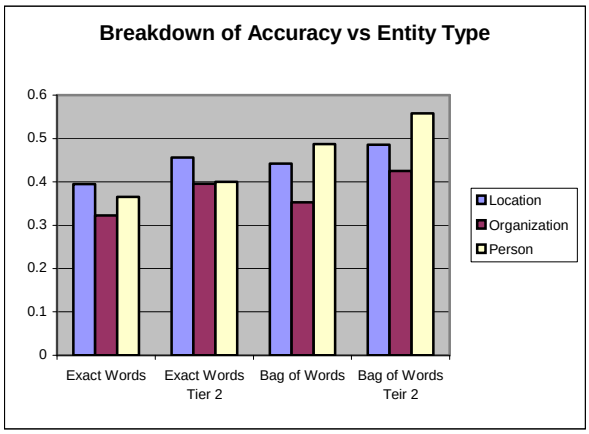

Figure 12. Breakdown of Accuracy for each evaluation category.

Conclusions and Discussion

We describe an implementation and evaluation of techniques for generating dynamic, query-driven summaries of links and the underlying media source streams with respect to named entities.

The Named Entity Connection graphs shows promise in the development of focused multi-story video summaries. The summaries generated automatically based on a limited database are similar to the human created summaries based on broad knowledge of the topics. This opens the door to future research for the application of link analysis for summarization of video databases. It also points to strengths and weaknesses in the underlying named entity detection. Future evaluations should test if merging of entities can increase accuracy due to mislabeled entities. While Infoplease is one benchmark for summaries, it would be useful to have a human evaluator judge the accuracy of the automatically generated named entity connection lists. There are many features that must be addressed when deciding how to improve the accuracy for this evaluation. Location names can be naively matched using exact phrasings because there are a limited number of ways to express a place name. Organizations and names of people can vary with abbreviations, acronyms and diminutives. This places a higher emphasis on the canonicalization of the entity names. A difficulty in directly comparing the Infoplease summaries and the automatically generated ones is the concept of scope. In the instance of Bill Clinton, the summary author used context from previous years to describe the person. The automatically generated summary via Named Entity Connections uses only the scope of the data given to it. For this evaluation, only 2004 data was given to the system, resulting in a summary of the entity for events in 2004. The usefulness of this distinction will be based on the needs of the user and can be investigated in the future. As evaluated, the system contains two implicit free variables, the maximum number distance the pairs of entities to use and the minimum threshold for links to be considered relevant. Additional experiments are required to determine the validity of choices used for those constants.

The evaluation was done on the 59 of the 93 possible names listed by Infoplease that had enough information in the database to create a named entity graph. The number in use is limited by three related factors: the result set size, the named entity detection and the scope due to the distance metric used by the connection process. Some of the entities listed by Infoplease had very little news coverage within the database used which caused the result set size to be very small. Entities that had fewer than four unique stories in their result set were disregarded automatically as an empirical cutoff. This removed some of the 93. While our technique for detecting named entities is good, it does not have 100% recall, which reduces the number of sentences given to the graph generation code. A graph is not considered for evaluation if the entire graph represents fewer than two stories. The distance metric used requires that two named entities be separated by at most one unrelated sentence. Due to the previous reasons, some potential pairings within the allowed distance could be missed. Also, some potential pairings are filtered out due to being too far apart.

Acknowledgments

This material is based on work supported by the Advanced Research and Development Activity (ARDA) under contract numbers H98230-04-C-0406 and NBCHC040037.

References

[1] Informedia Project web site at Carnegie Mellon University, http://www.informedia.cs.cmu.edu.

[2] Yan R., and Hauptmann, A. The Combination Limit of Video Retrieval. *Proc. ACM Multimedia (MM2003)*, Berkeley, CA.

[3] Hauptmann, A., and Witbrock, M.J., "Informedia: Newson-Demand Multimedia Information Acquisition and Retrieval", *Intelligent Multimedia Information Retrieval*, M. Maybury, Ed., AAAI Press/MIT Press, Menlo Park, CA, 1997, pp. 213-239.

[4] Cox, R.V., Haskell, B.G., Lecun, Y., Shahraray, B., and Rabiner, L., "Applications of Multimedia Processing to Communications", *Proc. of IEEE*, 1998, pp. 754-824.

[5] Christel, M., Warmack, A., Hauptmann, A., Crosby, S.,**.** Adjustable Filmstrips and Skims as Abstractions for a Digital Video Library**,** *IEEE Advances in Digital Libraries Conference 1999*, Baltimore, MD. pp. 98-104, May 19-21, 1999.

[6] Infoplease.com. 2004 People in the News, © 2005 Learning Network, http://www.infoplease.com/ipa/A092091.html.

[7] Christel, M., Hauptmann A., Wactlar, H., and Ng, T. Collages as Dynamic Summaries for News Video. *Proc. ACM MM '02* (Juan-les-Pins, France, December 2002), ACM Press.

[8] Merlino, A., Morey, D., and Maybury, M. Broadcast News Navigation using Story Segmentation, in *Proc. ACM Multimedia '97* (Seattle WA, Nov. 1997), ACM Press, 381-391.

[9] Bikel, D. M., Miller, S., Schwartz, R., and Weischedel, R. 1997. Nymble: a high-performance learning name-finder, in *Proc. 5th Conf. on Applied Natural Language Processing (ANLP)* (Washington DC, April 1997), 194-201.

[10] Miller, D., Schwartz, R., Weischedel, R., and Stone, R. Named Entity Extraction for Broadcast News, in *Proc. DARPA Broadcast News Workshop* (Washington DC, March 1999), http://www.nist.gov/speech/publications/darpa99/html/ie20/ie20.htm.

[11] Clarkson, P. and Rosenfeld, R. Statistical language modeling using the CMU-Cambridge toolkit, in *Proc. Eurospeech '97* (Rhodes, Greece, Sept. 1997), Int'l Speech Communication Assoc., 2707-2710.

[12] Ellson, J., Gansner, E., Kousofios, E. and North, S., Graphviz, http://www.graphviz.org

[13] Bagga A., Baldwin, B., Coreference as the Foundations for Link Analysis Over Free Text Databases. *Proc. COLING-ACL'98 Content Visualization and Intermediate Representations Workshop (CVIR '98)*, pp 19-24, August 1998.

[14] Hauck, R., Atabakhsh, H., Ongvasith, P., Gupta, H., and Chen, H., Using Coplink to Analyze Criminal-Justice Data, *Computer*, IEEE Computer Society Press, Lot Alamitos, CA, USA, March 2002.

[15] Hong, T., and Han, I. Knowledge-based Data Mining of News Information on the Internet Using Cognitive Maps and Neural Networks. *Expert Systems with Applications*, 23, 1–8. 2002.

[16] Huang, X. D., Alleva, F., Hwang, M.-Y., and Rosenfeld, R, An Overview of the SPHINX-II Speech Recognition System. In *Proceedings of the ARPA Human Language Technology Workshop*, published as *Human Language Technology*, pages 81–86. Morgan Kaufmann, March 1993.

[17] Sato, T., Kanade, T., Hughes, E., and Smith, M., "Video OCR for Digital News Archive," *Proc. Workshop on Content-Based Access of Image and Video Databases*, IEEE CS Press, Los Alamitos, Calif., 1998, pp. 52-60.