Event Information Extraction System (EIEE): FSM vs HMM

Shaukat Wasi, Zubair A. Shaikh, Sajid Qasmi, Hussain Sachwani, Rehman Lalani, and Aamir Chagani

***Abstract*—**Automatic Extraction of Event information from social text stream (emails, social network sites, blogs etc) is a vital requirement for many applications like Event Planning and Management systems and security applications. The key information components needed from Event related text are Event title, location, participants, date and time. Emails have very unique distinctions over other social text streams from the perspective of layout and format and conversation style and are the most commonly used communication channel for broadcasting and planning events. Therefore we have chosen emails as our dataset. In our work, we have employed two statistical NLP methods, named as Finite State Machines (FSM) and Hidden Markov Model (HMM) for the extraction of event related contextual information. An application has been developed providing a comparison among the two methods over the event extraction task. It comprises of two modules, one for each method, and works for both bulk as well as direct user input. The results are evaluated using Precision, Recall and F-Score. Experiments show that both methods produce high performance and accuracy, however HMM was good enough over Title extraction and FSM proved to be better for Venue, Date, and time.

***Keywords*—**Emails, Event Extraction, Event Detection, Finite state machines, Hidden Markov Model.

I. Introduction

A LOT of applications really need event related information to be extracted and being available on hand. Security applications require such information to detect frauds etc. Personal management systems need event information to support event management and planning tasks. Even, an application that automatically extracts event information and posts it to one's calendar would be highly appreciated by the users. Extracting event information means the extraction of date, time, venue, participants and the title of the event.

Extracting Event related information from emails is much trickier than extraction from some other text stream. The difference lies in the syntax and format of the text and the wide range of event types. The conversation styles normally provide clues about the event category. Emails normally have short descriptions and therefore may not contain enough description of the event. Secondly, since emails are not only used for formal event announcements, but these are also widely used for event planning and discussions. Therefore, emails contain variations on layout and text organization. An invitation to a personal event will normally be of a free style with no formalities over layout and text. Instead, an official event announcement will be placed confined in a specific format and layout with proper greetings and signoff texts.

We have developed a system that uses FSM and HMM to extract the event information components from the emails. The objective over using two different techniques was to compare the efficiency and accuracy of the two techniques with respect to the extraction task over event emails. A finite state machine (FSM) is a mathematical model. It can be thought of an abstract machine that has finite number of states. The machine can be in any one state at a particular time and the movement from one state to other is called transition. FSM has one initial state and a finite number of final states. Hidden Markov Model is also a statistical model that is represented through a number of states and transitions. Some of the states are hidden, that is they are to be predicted through some process.

The rest of the paper is organized as follows. Section II discusses the related work. Section III explains event information extraction system. Section IV provides the results of evaluation and finally Section V has the conclusion followed by references in section VI.

II. Related Work

Several attempts have been made over the event extraction task with different objectives and perspectives. Event extraction from general text streams (news wires, articles etc.) [1-8] was mostly aimed towards the detection of hot or new events. Most of the work in general text streams did not use the actual time of event occurrence and just focused the publication date.

Event information extraction from social text streams has been targeted by many researchers in the recent years. Some authors [9-11] have taken event as a topic with continuous discussion over a specific period of time. They have provided dimensional visualizations of the events with respect to time, location and people and have extracted this information by observing the conversations over time line and using a bag of

Shaukat Wasi is with the National University of Computer and Emerging Sciences (FAST-NU), ST-4, Sector 17-D, Shah Latif Town, Karachi, PAK (phone: 92-21-34100541; e-mail: shaukat.wasi@ nu.edu.pk).

Zubair Shaikh is with the National University of Computer and Emerging Sciences (FAST-NU), ST-4, Sector 17-D, Shah Latif Town, Karachi, PAK (phone: 92-21-34100541; e-mail: zubair.shaikh@ nu.edu.pk).

Sajid Qasmi was with the National University of Computer and Emerging Sciences (FAST-NU), ST-4, Sector 17-D, Shah Latif Town, Karachi, PAK (phone: 92-21-34100541; e-mail: sajid.qasmi@gmail.com).

Hussain Sachwani is with the National University of Computer and Emerging Sciences (FAST-NU), ST-4, Sector 17-D, Shah Latif Town, Karachi, PAK (phone: 92-21-34100541; e-mail: k102097@ nu.edu.pk).

Rahman Lalani is with the National University of Computer and Emerging Sciences (FAST-NU), ST-4, Sector 17-D, Shah Latif Town, Karachi, PAK (phone: 92-21-34100541; e-mail: k102006@ nu.edu.pk).

Amir Chagani is with the National University of Computer and Emerging Sciences (FAST-NU), ST-4, Sector 17-D, Shah Latif Town, Karachi, PAK (phone: 92-21-34100541; e-mail: k102138@ nu.edu.pk).

words type approach. An attempt to identify announcement through emails [12] used words tagged with Named Entities. They applied Naïve Bayes Classifier to extract the required information. A statistical model was proposed by Lin et al[13] to detect the events that are popular in a social community. Experiments with PET were performed over DBLP and Twitter. A work on weblogs[14] targeted the events appearing in news that were discussed over the weblogs. A relevance model based frame work was provided to track such events. A keyword based approach is also proposed for detecting and tracking new events[15]. A key word graph approach was used by Hassan Sayyadi[15]to detect and track event s in social text streams. H. Becker et al [16, 17]incorporated the tags and labels, available within the discussion over events in social media sites, to identify the events.

There are works that used Finite State Machines (FSM) and Hidden Markov Models (HMM) for extracting specific event related information meant for specific purposes. FSM were used to extract temporal information from Korean news articles [18]. The objective was to draw the attention, of the research community working in the event detection area, towards the importance of extracting actual temporal information available within the text. A system named FASTUS[19] used FSM for the message understanding task and proved that FSM is an efficient method as compare to other natural language processing techniques. Another research work [20] that employed FSM for extracting event related information constructed distinct sets of FSMs for extracting phrases that reveal location, time and title information. We have modified the FSMs provided by [20] for further improvement towards precision and recall. An information extraction system, developed for mobile devices [21] used HMM for extracting location and topic and used FSM for extracting temporal information. The system was tested over Korean text messages.

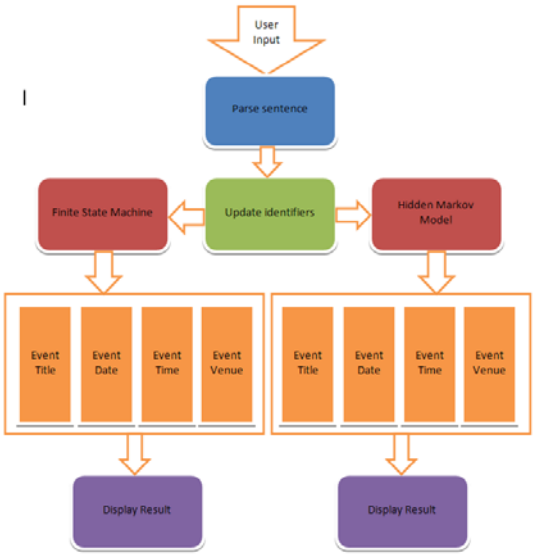

Fig. 1 Architecture of EIES

III. Event Information Extraction System

Our system comprises of two modules, one that employs FSM and the other using HMM for the event information extraction task. Using C#, we have designed two separate APIs for FSM and HMM. Both have been interfaced with a windows forms based application which inputs the sentence and extracts the required information with the help of these APIs. SQL Server has been used as the backend database. The application takes input in two ways. The user may enter text directly into a text box or he may upload a text file for bulk input. The architecture of the system is provided in Fig. 1 and Fig. 2 shows the snapshots of the application. The details regarding the two methods (FSM and HMM) are provided in the next sub sections. In both cases, the transitions between states are over parts of speech (POS). This is a unique distinction over most of the extraction techniques that use actual words from the text. The use of POS provides a generalization towards the textual representation of the event, means it supports the handling of variations in conversation layout and format. The rules under both methods are simply based on the grammatical rules of English or we can simply say that these rules are the mathematical translation of Englishgrammar rules. The rules are applicable only for the email dataset. The reason for addressing only email data set is described, to some extent, in the related work section.

Actually, there are a lot of techniques available for other datasets and much effort has been made towards these data sets (news articles, scientific documents etc). Secondly the objectives under the extraction task over emails are dissimilar to the extraction over other data sets. We aim to support event planning and management tasks through automatic extraction of event related information.

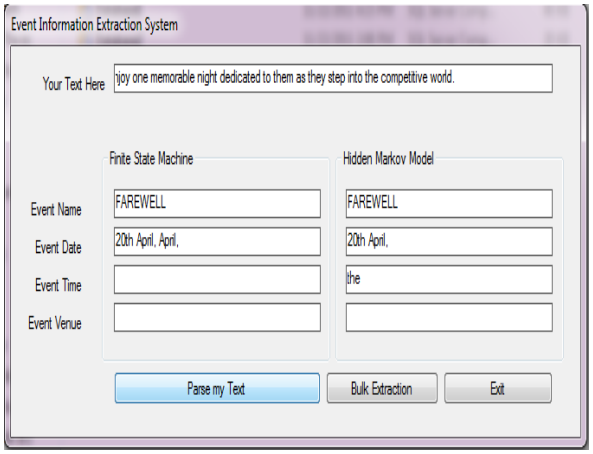

Fig. 2 Main interface of EIES that provides options for direct input as well as bulk input

A. Extracting Event Information using Hidden Markov Model

We have slightly modified the HMM for our purpose. Extraction of each information component of an event description has been tackled independently. The set of rules for each case are defined below. We have customized it in the

best extent to get more accurate information. The list of identifiers used in describing the states of the HMM are given in Table I.

1. HMM Rules for Event Title Extraction:

Different cases for extraction of titles using HMM are discussed below one by one.

TABLE I
SET OF IDENTIFIERS USED TO DEFINE HMM STATES

Identifier	Definition
<AT>	Articles
<PA>	Adverbial Particles
<VBX>	Auxiliary Verbs
<PX>	Auxiliary Verbs
<NNBU>	Bound Noun for Units
<NNPB>	Bound Proper Nouns
<BS>	Blank State
<NX>	Proper Noun
<NNPD>	Demonstrative Pronoun
Sn	Input State
NF	Result Not Found
O	Desired Output

Case 1:HMM searches for an article and changes its state to a blank state, and now it search for an adverbial particle if it cannot find any adverbial particle it moves to Result not found (NF) state, if adverbial particle is found, HMM searches for an auxiliary verb and moves to its desired output (O) state. If auxiliary verb is not found it moves to NF state rather than moving towards O state. The following sentence is an example depicting the type of phrases identified by HMM. "DECS has invited FASTIANS on his 2012 Annual Dinner." Fig. 3 shows the state diagram for this case.

Case 2:HMM searches for an article and changes its state when an article is searched it goes for a blank state. After a blank state it moves for an adverbial particle and auxiliary verb and finally reaches to its desired output (O) state. If adverbial particle, auxiliary verb or blank state is not found it moves toward Result not found (NF) state. Example for the invitation phrases handled in this case is "Batch 08 have been requested by SIBBS to attend workshop conducted by NOKIA". Fig. 4 shows the state diagram for case 2 of title detection.

Case 3: HMM searches for an adverbial particle and changes its state to a blank state. After a blank state it searches for an auxiliary verb and adverbial particle respectively, if any one of the verb is not found it moves toward Result not found (NF) state otherwise it moves to a blank state, and again it searches for auxiliary verb and after that it reaches to its desired output (O) state. If auxiliary verb is not found it moves toward Result not found (NF) state. An example which describes the case is "a workshop on android developing to be held". Fig. 5 shows the state diagram for this case.

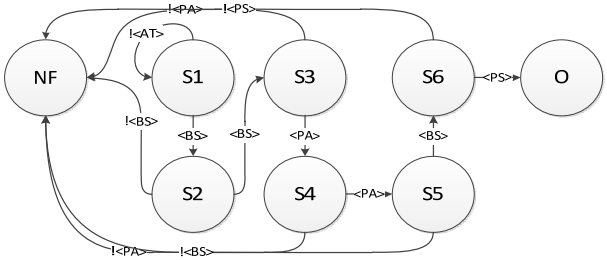

Fig. 3 HMM Event Title Extraction Case 1

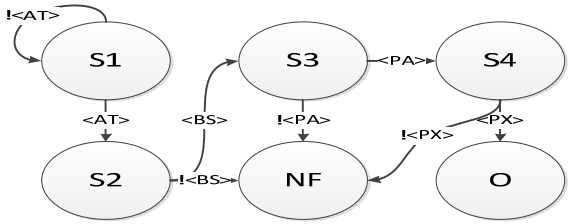

Fig. 4 HMM Event Title Extraction Case 2

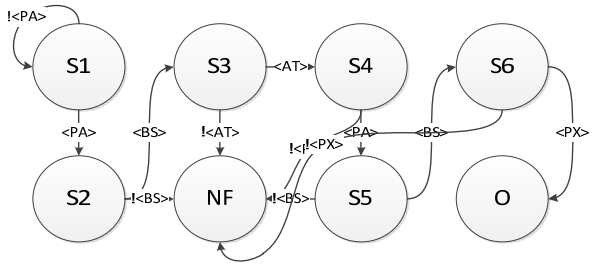

Fig. 5 HMM Event Title Extraction Case 3

2. HMM Rules for Event Time Extraction

Different cases for extraction of time using HMM are discussed below one by one.

Case 1: In HHM rules for Event Time Extraction case-1, HMM searches for an auxiliary verb and adverbial particle respectively and moves toward a blank state and after that it reaches to its desired output (O) state, if adverbial particle or a blank state is not found it moves toward Result not found (NF) state. Example which is describing the phrases in case of event time extraction is "please guys meet me at 4:00 pm in R12". Fig. 6 shows the state diagram for case 1 of Time extraction.

Case 2: In HHM rules for Event Time Extraction case-2, HMM searches for an auxiliary verb, preposition, adverbial particle and again searches for preposition respectively and moves toward a its Desired output (O) state, if adverbial particle or a preposition are not found it moves toward Result not found (NF) state. Here is an example for the case of event time extraction "seminar will be conducted by Ali in auditorium timings are from 4:00 pm to 4:30 pm". Fig. 7 shows the state diagram for case 2 of Time extraction.

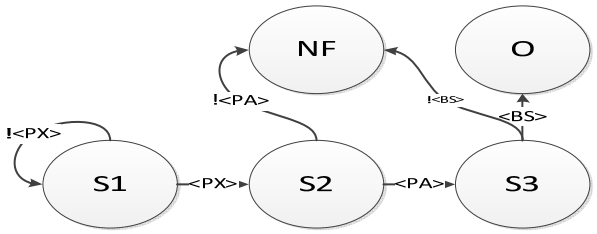

Fig. 6 HMM Event Time Extraction Case 1

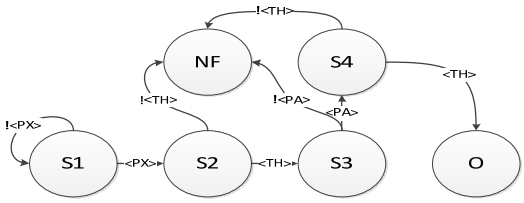

Fig. 7 HMM Event Time Extraction Case 2

3. HMM Rules for Event Date Extraction

Different cases for extraction of date using HMM arediscussed below one by one.

Case 1: In this case, HMM searches for an adverbial particle and an article respectively and moves toward its desired output (O) state, if an article is not found it reaches toward Result not found (NF) state. Example which classifies the event date extraction is "examination papers are approved by higher authorities Dated 13th November 2012". The state diagram for the case is shown in Fig. 8.

Case 2: HMM searches for bound proper nouns, adverbial particle and article respectively. If adverbial particle or article is not found it moves to Result not found (NF) state or else it moves to next state and finally reaches to its desired output (O) state. Fig. 9 contains the state diagram for this case. An example that describes the phrase in event date extraction is "remaining classes are schedule on Tuesday 13th November 2012".

4. HMM Rules for Event Venue Extraction

Different cases for extraction of venue using HMM are discussed below one by one.

Case 1: In HHM rules for Event Venue Extraction case-1, HMM searches for a proper noun, after that it moves for a blank state and finally reaches to its desired output (O) state. If it does not get any blank state it moves toward Result not found (NF) state. An example for event venue extraction is "With its ideal location at the heart of Chicago downtown and close to the Loop, the Holiday Inn Mart Plaza ChicagoHotel". The state diagram for this case is given in Fig. 10.

Case 2: In HHM rules for Event Venue Extraction case-2, HMM searches for a proper noun and article respectively, after that it moves to a blank state and finally reaches to its desired output (O) state. If it does not get any blank state or an article it moves toward Result not found (NF) state. The state diagram for this case is given in Fig. 11. Given blow is an example for event venue extraction case "Location at FAST-NUCES. We will meet after second slot for PROCOM.NET".

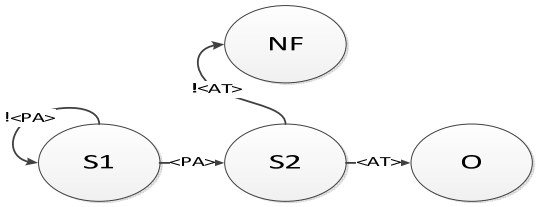

Fig. 8 HMM Event Date extraction Case 1

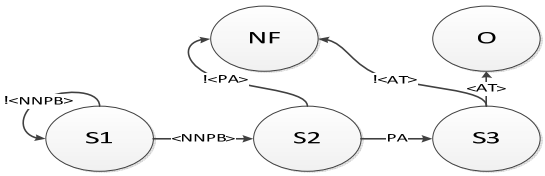

Fig. 9 HMM Event Date extraction Case 2

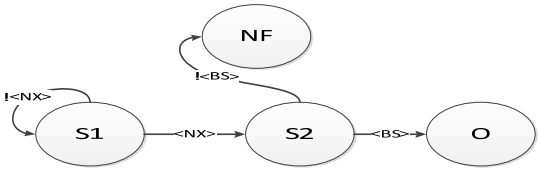

Fig. 10 HMM Event Venue extraction Case 1

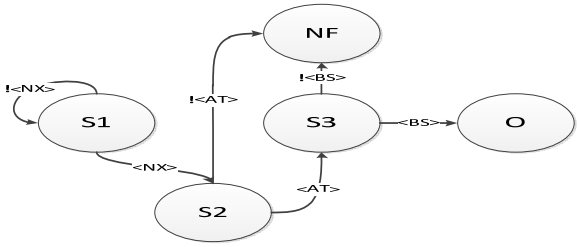

Fig. 11 HMM Event Venue extraction Case 2

B. Extracting Event Information using Finite State Machine

Another technique used for the extraction of information was FSM. Like HMM, independent cases have been defined for extraction of individual event related information components.

1. FSM Rules for Event Title Extraction

Different cases for extraction of title using FSM are discussed below one by one.

Case 1: This FSM will be accepted, when it will get an article like "a, an, the" and then "noun" in the sentence. If in the sentence if it will not find an article then it will remains in state 1, after when it will find an article then it will further move to state 2, then if it will not find a noun then it will move to NF State (Not Final) and when it will find a noun then it will move to final accepting state. The following sentence is an example depicting the type of phrases identified by FSM. "You are requested to attend series of seminar on software engineering conducted by various software houses." Fig. 12 contains the FSM for this case.

Case 2: This FSM will be accepted, when it will get an "article" like (a, an, the), then "Adverbial Particle" like (at, in, on) and then "noun" in the sentence. If in the sentence if it will

not find an article then it will remains in state 1, after when it will find an article then it will further move to state 2, then if it will not find an Adverbial Particle then it will move to NF State (Not Final) and when it will find a Adverbial Particle then it will move to State3, then if it will not find a noun then it will move to NF State (Not Final) and when it will find a noun then it will move to final accepting state. The following sentence is an example depicting the type of phrases identified by FSM. "Please attend my birthday party; we will meet in KFC Garden West outlet." The FSM for this case is provided in Fig. 13.

Case 3: This FSM will be accepted, when it will get an "article" like (a, an, the), then "Auxiliary" (From, to, for), then "Adverbial Particle" like (at, in, on) and then "noun" in the sentence. If in the sentence if it will not find an article then it will remains in state 1, after when it will find an article then it will further move to state 2, then if it will not find an Auxiliary then it will move to NF State (Not Final) and when it will find a Auxiliary then it will move to State3, then if it will not find an Adverbial Particle then it will move to NF State (Not Final) and when it will find a Adverbial Particle then it will move to State4, then if it will not find a noun then it will move to NF State (Not Final) and when it will find a noun then it will move to final accepting state. Fig. 14 shows the state diagram for the case. The following sentence is an example depicting the type of phrases identified by FSM. "Please join us on marriage ceremony of my son Ali."

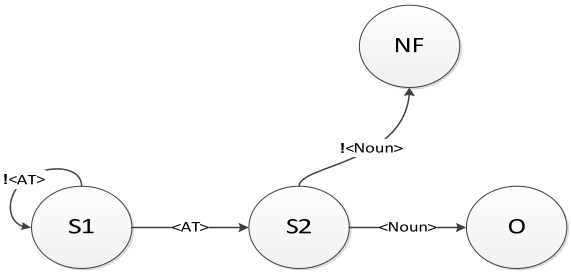

Fig. 12 FSM Event Title extraction Case 1

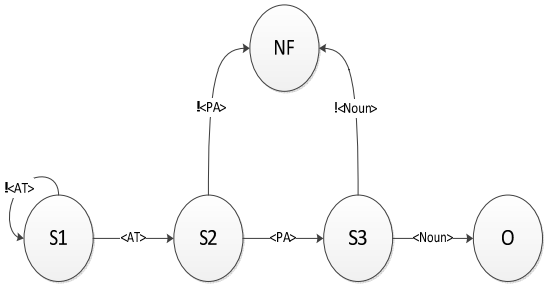

Fig. 13 FSM Event Title extraction Case 2

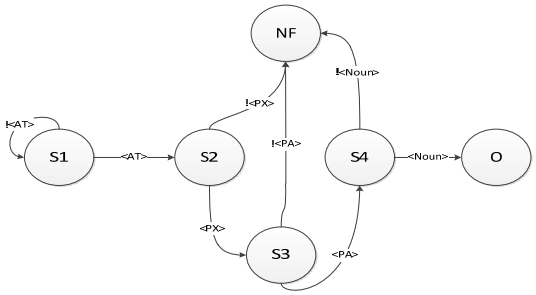

Fig. 14 FSM Event Title extraction Case 3

2. FSM Rules for Event Time Extraction

Different cases for extraction of time using FSM are discussed below one by one.

Case 1: This FSM will be accepted, when it will get an "auxiliary" like (From, to, for), then "TM" and then "TH" in the sentence. If in the sentence if it will not find an auxiliary then it will remains in state 1, after when it will find an auxiliary then it will further move to state 2, then if it will not find a TM then it will move to NF State (Not Final) and when it will find a TM then it will move to State3, then if it will not find a TH then it will move to NF State (Not Final) and when it will find a TH then it will move to final accepting state. The FSM for this case is given in Fig. 15. The following sentence is an example depicting the type of phrases identified by FSM. "We will meet at 3:30 PM for DECS society meeting".

Case 2: This FSM will be accepted, when it will get an "auxiliary" like (From, to, for), then "TM" and then "TH" in the sentence. If in the sentence if it will not find an auxiliary then it will remains in state 1, after when it will find an auxiliary then it will further move to state 2, then if it will not find a TM then it will move to NF State (Not Final) and when it will find a TM then it will move to State3, then if it will not find a TH then it will move to NF State (Not Final) and when it will find a TH then it will move to state4, then if it will not find an auxiliary then it will move to NF State (Not Final), after when it will find an auxiliary then it will further move to state5, then if it will not find a TM then it will move to NF State (Not Final) and when it will find a TM then it will move to State6, then if it will not find a TH then it will move to NF State (Not Final) and when it will find a TH then it will move to final accepting state. The following sentence is an example depicting the type of phrases identified by FSM. "Probably we will gather in KFC between 5:30 PM and 8:00 PM." FSM for the case is contained in Fig. 16.

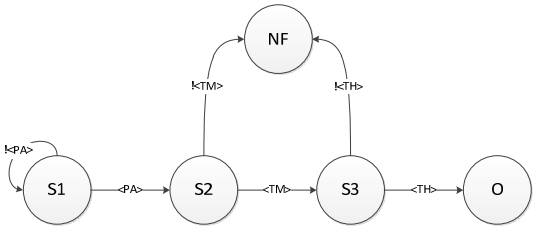

Fig. 15 FSM Event Time extraction Case 1

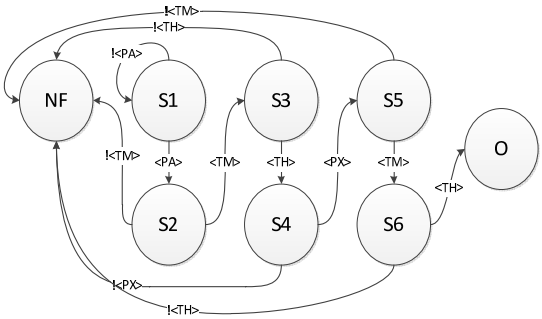

Fig. 16 FSM Event Time extraction Case 2

3. FSM Rules for Event Date Extraction

Different cases for extraction of date using FSM are discussed below one by one.

Case 1: In this case first FSM will search for an auxiliary particle if an auxiliary particle is found then it will move to the next state otherwise it will remain on the same state. On the next state it will search for (NNPB) bound proper nouns. If NNPB will found then FSM will give the desired output else it will move to the NF state. Fig. 17 shows the FSM used for this case. The following sentence is an example depicting the type of phrases identified by FSM. "Examination papers are approved by higher authorities Dated 13th November 2012".

Case 2: In this case FSM will first search for (NNBU) bound noun for units if found then it will move to the next state (S2) else it will remain on the same state (S1). On the next state (S2) it will search for (NNPB) bound proper nouns. If NNPB will found then FSM will give the desired output else it will move to the NF state. The following sentence is an example depicting the type of phrases identified by FSM. "Tomorrow 15th April 2012: Annual Dinner." FSM for the case is shown in Fig. 18.

Case 3: In this case FSM will first search for (NNPB) bound proper nouns if it is found then FSM will move to the next state otherwise it will remain on the same state. On the next state it will again search for an (NNPB) if it is found then it will give the output otherwise it will move to the NF state. FSM for the case is shown in Fig. 19. The following sentence is an example depicting the type of phrases identified by FSM. "We friends have a get to gather in Rehmat-e-Shireen on Saturday 21st April 2012".

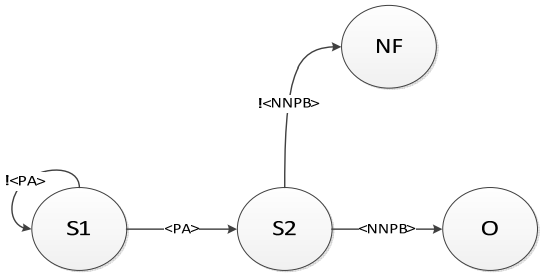

Fig. 17 FSM Event Date extraction Case 1

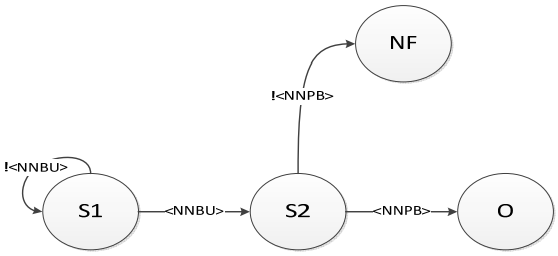

Fig. 18 FSM Event Date extraction Case 2

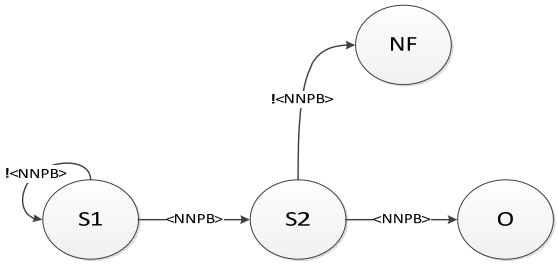

Fig. 19 FSM Event Date extraction Case 3

4. FSM Rules for Event Venue Extraction

Different cases for extraction of venue using FSM are discussed below one by one.

Case 1: In FSM rule for event venue extraction first FSM will search for an (PA) auxiliary article if an auxiliary article is found it will move to the next state otherwise it will stay on the same state. On the next state it will search (PB) proper noun if PB not found it will go to NF State otherwise it will move to the next state. On the next state (S3) it will search for the noun if found it will give the desired output otherwise move to the NF State. Fig. 20 contains the FSM for this case. Example to extract venue "Event is organized by SIBBS and Located on Karachi Hall".

Case 2: In this case first FSM will search for an auxiliary article if an auxiliary article (PA) found it will move to the next state otherwise remain on the same state. On the second state it will search for noun. If a noun found then it will goes to the final state and give an output else if noun not found then it will goes to the NF State. Example for event venue extraction is "marriage ceremony venue is Karachi Hall". Fig. 21 shows the FSM for case 2 of Event Venue extraction through FSM.

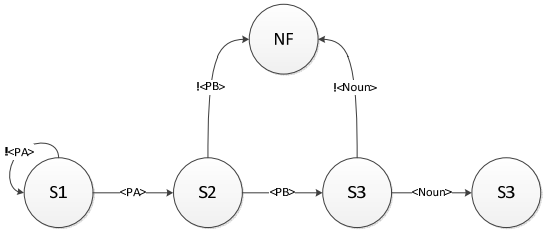

Fig. 20 FSM Event Venue extraction Case 1

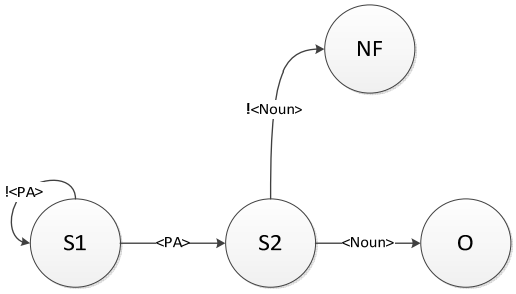

Fig. 21 FSM Event Venue extraction Case 2

IV. Results

We have tested both techniques, FSM and HMM, over 3647 mails collected from our inboxes. Out of these 3647 emails, 2910 contained real events and the rest 737 were non-event emails. Although our system supports two methods for providing testing data, we have used the bulk input option for testing the system. Fig. 22 and Fig. 23 contain the precision, recall and F1 score obtained through FSM and HMM respectively. The graph in Fig. 24 shows the comparison of both techniques, over extraction of individual event information components. The results show that FSM provide higher precision and recall over location and time and HMM produces better results for event title extraction. Higher precision over location attribute means that most of the retrieved results that are marked as location are correctly identified by the system. The same meaning goes over all the event related attributes. Similarly, high recall over location attribute means that most of the locations have been retrieved by the system i.e. very few are missed. So, through FSM, high precision and recall over location, time, and date means that FSM retrieves most of the attributes available in the emails' text and most of the information retrieved is relevant. Similarly in case of HMM, it was successful in retrieving the title attribute from emails and more than 90% of the titles extracted were correct.

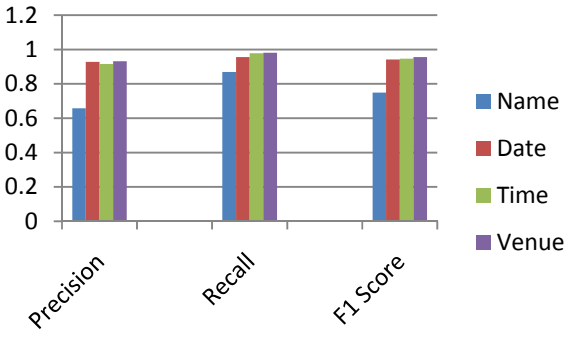

Fig. 22 Precision, Recall, and F1 score obtained through FSM

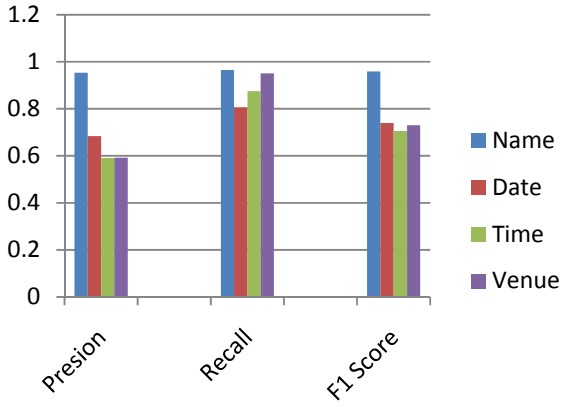

Fig. 23 Precision, Recall, and F1 score obtained through HMM

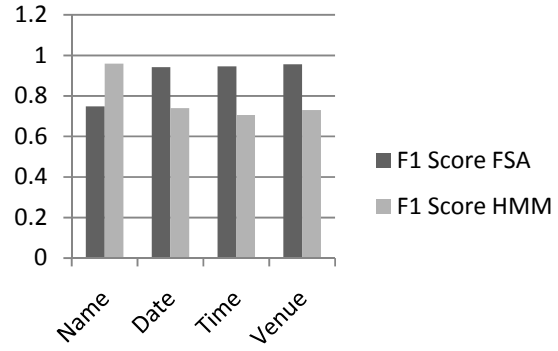

Fig. 24 F1 score – HMM VS FSM

V. Conclusion

Through the implementation and comparison of results obtained from Finite State Machine and Hidden Markov Model it has been observed that FSM provides more accurate information regarding time, date and venue and it is weaker in extraction of event title. However, HMM is more powerful when it comes to the extraction of event title and provides up to more accurate results and correct information, but it lacks in extracting event time, date and venue due to its probabilistic nature.

By comparing both the techniques it is now clear that a more customized HMM can provide more accurate information as compared to FSM. And the combination of both will be smart enough to produce a precision and recall values closer to 1. It means that most of the information will be retrieved and more than 90% of the retrieved attributes will be correct and relevant.

Acknowledgment

This research has been supported by HEC, Pakistan. We pay thanks to the CRUC (Center for research in Ubiquitous computing) team at FAST-NU Karachi that provided us the environment promoting such research activities.

REFERENCES

[1] J. Allan, et al., "Topic Detection and Tracking Pilot Study Final Report," in DARPA Broadcast News Transcription and Understanding Workshop, 1998.

[2] J. Allan, R. Papka, and V. Lavrenko, "On-Line New Event Detection and Tracking," presented at SIGIR'98, Melbourne, Australia, 1998.

[3] Y. Yang, T. Pierce, and J. Carbonell, "A Study on Retrospective and Online Event Detection," presented at SIGIR'98, Melbourne, Australia, 1998.

[4] Y. Yang, et al., "Learning Approaches for Detecting and Tracking News Events," IEEE Intelligent Systems Special Issue on Applications of Intelligent Information Retrieval, vol. 4, pp. 32-43, 1999.

[5] G. Kumaran and J. Allan, "Text Classification and Named Entities for New Event Detection," presented at SIGIR'04, Sheffield, South Yorkshire, UK, 2004.

[6] D. Kusui, K. Tateishi, and T. fukushima, "Information Extraction and Visualization Fro Internet Documents," NEC Journal of Advanced Technology, vol. 2, 2005.

[7] K. Chen, L. Luesukprasert, and S. T. Chou, "Hot Topic Extraction Based on Timeline Analysis and Multidimensional Sentence Modeling," IEEE Transactions on Knowledge and Data Engineering, vol. 19, 2007.

[8] Z. Kuo, L. J. Zi, and W. Gang, "New Event Detection Based on Indexing-Tree and Named Entity," presented at SIGIR'07, Amsterdam, The Netherlands, 2007.

[9] X. Wan, E. Milios, and N. Kalyaniwalla, "Link-Based Event Detection in Email Communication Networks," presented at SAC'09, Honolulu, Hawaii, U.S.A, 2009.

[10] Q. Zhao and P. Mitra, "Event Detection and Visualization for Social Text Streams," presented at ICWSM, Coloroda, USA, 2007.

[11] Q. Zhao, P. Mitra, and B. Chen, "Temporal and Information Flow Based Event Detection from Social Text Streams," presented at American Association for Artificial Intelligence (AAAI 2007), Vancouver, British Columbia, Canada 2007.

[12] V. Pekar, "Information Extraction from Email Announcements," in Lncs, Natural Language Processing and Information Systems. Berlin Heidelberg: Springer Verlag, 2005, pp. 372-375.

[13] C. X. Lin, et al., "Pet: A Statistical Model for Popular Events Tracking in Social Communities," presented at SIGKDD, New York, USA, 2010.

[14] V. Ha-Thuc, et al., "Event Intensity Tracking in Weblog Collections," presented at ICWSM-DCW' 09, California, USA, 2009.

[15] H. Sayyadi, M. Hurst, and A. Maykov, "Event Detection and Tracking in Social Streams," presented at Association for Advancement of Artificial Intelligence (AAAI'09), 2009.

[16] H. Becker, M. Naaman, and L. Gravano, "Event Identfication in Social Media," presented at Twelfth International Workshop on the Web Databases (WebDB 2009), Providence, USA, 2009.

[17] H. Becker, M. Naaman, and L. Gravano, "Learning Similarity Metrics for Event Identification in Social Media," presented at WSDM, New York, USA, 2010.

[18] P. King and S. H. Mayeng, "Usefulness of Temporal Information Automatically Extracted from News Articles for Topic Tracking," ACM Transactions on Asian Language Information Processing, vol. 3, pp. 227-242, 2004.

[19] J. HOBBS, et al., "Fastus: Acascaded Finite-State Transducer for Extracting Information from Natural-Language Text," presented at MUC, Cambridge, MA, 1997.

[20] S. Wasi, Z. Shaikh, and J. Shamsi, "Contextual Event Information Extractor for Emails," SURJ, 2011.

[21] C.-N. Seon, H. Kim, and H. Kim, "Information Extraction Using Finite State Automata and Syllable N-Gramsin a Mobile Environment," presented at ACL-08: HLT Workshop on Mobile Language Processing, Ohio, USA, 2008.