5. Features for classification

Lexical and syntactic features of tokens and entity pairs that created prior to classification are used to build the classification model. These features are a part of those described in [41] and [42]. These features are split into 14 sets as described in table 2.

TokN features are contained surface string and POS of the tokens that surrounding the entity pairs. This features are provide us with important information about the words surrounding entity pairs to decide if there is relationship between them. GentokN features are generalised tokN which containing morphological root and generalised POS. Str features are contained surface string features include all token features of both entity pairs, their heads, combine of their heads, first, last and other tokens between them, two tokens before the leftmost and after the rightmost entity pairs. POS features are created from POS tags of the entity pairs and the tokens that surrounding them. Root features are created from morphological analyzer of the entity pairs and the tokens that surrounding them. GenPOS features are created from generalised POS tags of the entity pairs and the tokens that surrounding them. Entities were divided into two categories are events and non-events entities. Event entities are Investigation and Intervention entities. Non-event entities are Condition, Location, Drug-device, Result, Negation, Laterality, and Sub-location. Inter features are contained intervening entities which mean types and numbers of entities between entity pairs. Event features are contained whether an entity pairs contain two events, two non-events, or one event and one non-event and if there are any intervening events or non-events between entity pairs. Allgen features are collection of all above features in root and generalised POS forms. Notok features are collection of all above features except for TokN.

Stanford Parser [43] can be applied to parse the corpus to generate a dependency analysis which contains syntactic relations between sentence tokens for the dep and syndist features sets. When the entities exist in the same sentence the dep feature set can be generated from the parse. This feature set consists of the dependency analysis of entity pairs, their heads, and combine of their heads, first, last and other tokens between them, two tokens before the leftmost and after the rightmost entity pairs. For the syndist feature set contains the number of links on the dependency path between the entity pairs and the number of tokens between two entities [28].

6. Results and Discussion

Evaluation of the system can be done by using the standard evaluation metrics of Recall and Precision. The terms of true positive (TP), false positive (FP) and false negative (FN) are used to determine Recall and Precision which matches between relations recorded in a system annotated response document and a gold standard key document. If the relation in the response exists in the key with the same arguments then the response relation is a true positive. If the relation in the response dose not exists in the key then the response relation is a false positive. If the relation in the key dose not exists in the response then the key relation is a false negative.

$$R=\frac{TP}{TP+FN} \qquad P=\frac{TP}{TP+FP} \qquad F1=\frac{2PR}{P+R}$$

Table 2: Feature sets for learning

Feature set	***Description***
TokN	Surface string and POS of tokens surrounding the arguments, windowed -*N* to +*N*, *N* = 6 by default.
GentokN	Root and generalised POS of tokens surrounding the argument entities, windowed *N* to +*N*, *N* = 6 by default.
Atype	Concatenated semantic type of arguments, in arg1-arg2 order.
Dir	Direction: linear text order of the arguments (is arg1 before arg2, or vice versa?).
Str	Surface string features based on Zhou et al [29], see text for full description.
POS	POS features, as above.
Root	Root features, as above.
GenPOS	Generalised POS features, as above.
Inter	Intervening mentions: numbers and types of intervening entity mentions between arguments.
Event	Events: are any of the arguments, or intervening entities, events?
Allgen	All above features in root and generalised POS forms, i.e. gen-tok6+atype+dir+root+genpos+inter+event.
Notok	All above except tokN features, others in string and POS forms, i.e. atype+dir+str+pos+inter+event
Dep	Features based on a syntactic dependency path.
Syndist	The distance between the two arguments, along a token path and along a syntactic dependency path.

Standard ten-fold cross validation methodology is used to split the corpus for evaluation in our experiments. There are scores for each type of relations and for relation overall. P, R and F1 scores are computed for each relation type on each fold and macro-averaging these values for individual relations.

6.1 Algorithm type

Multi algorithms are implemented on the training corpus of patient narratives to see which one is much suitable for relation extraction. Many algorithms of supervised machine learning are applied such as Naïve Bayes Weka, C4.5 Weka**,** KNN Weka, Perceptron algorithm with uneven margin (PAUM), and Support victor machine with uneven margin (SVM), the results of these algorithms are described in table 3.

Table 3: Relation extraction by different algorithms

Relationship type	***Metric (%)***	***Naive Bayes Weka***	***C4.5Weka***	***KNN Weka***	***PAUM***	***SVM UM***
Has_finding	P	48.48	0.42	70.76	0.67.5	76.66
	R	72.11	31.16	53.11	62.04	55.85
	F1	52.88	29.27	52.95	58.93	57.32
Has_indication	P	59.85	58.67	59.77	61.71	67.17
	R	80.39	94.79	84.76	85.59	72.85
	F1	67.57	71.43	68.75	70.6	68.89
Has_location	P	67.04	66.77	69.02	71.32	76.68
	R	94.98	95.63	91.73	92.28	85.67
	F1	78.12	78.2	78.28	79.7	80.11
Has_target	P	54.27	50.97	54.04	63.29	68.92
	R	95.39	96.65	86.16	86.98	76.45
	F1	68.71	66.36	66.01	72.88	71.59
Laterality_modifies	P	43.4	43.4	54.07	41.73	60
	R	58.57	58.57	68.57	58.57	51.9
	F1	47.85	47.85	58.57	47.52	54.23
Negation_modifies	P	62.44	72.38	71.11	70	80
	R	74.16	74.16	77.5	80	71.66
	F1	65.73	71.84	72.72	73.75	74.66
Sub-location_modifies	P	77.6	90.22	92.22	90.22	100
	R	98	98	98	98	93
	F1	85.24	93.14	94.03	93.14	95.55
Overall	P	60.95	60.81	63.33	67.34	73.99
	R	90.51	92.18	87.63	88.59	79.67
	F1	72.48	73.01	73.27	76.14	76.3
Run Time in seconds		28.563	29.999	25.148	18.736	27.487

Firstly; Naïve Bayes Weka algorithm is implemented. Different algorithm C4.5 decision tree is applied; overall F1value increases by around 0.5% than the value of NaiveBayesWeka algorithm. KNN Weka algorithm is used with the option '–k 2' to get the best results, there is small increase in the value of overall F1 around by 0.26% than the value of C4.5Weka. Another algorithm PAUM is implemented with the best options "–p 20 –n 5 –optB 0.0". Overall F1 value of PAUM improves the performance than the overall F1value of KNN Weka by around 3%. Finally; SVM with uneven margin algorithm is executed with the options "-c 0.7 -t 1 -d 2 -m 100 -tau 0.8" to get the best results. This means that polynomial kernel is used with degree 2 for quadratic kernel and parameter of uneven margin (τ) is 0.8. There is small change in the overall F1 value of SVM algorithm than overall F1 value of PAUM algorithm around by 0.16%.

From the results in table 3; SVM algorithm with uneven margin is the much suitable machine learning algorithm for relation extraction from medical texts. Figure 8 shows the graph of applying different algorithms for relation extraction from medical texts.

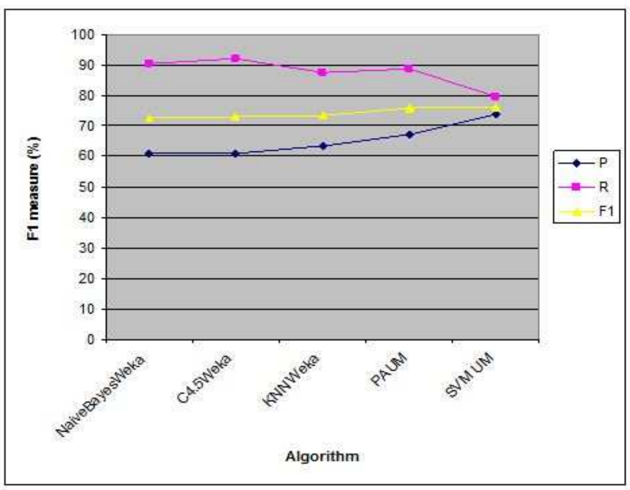

Fig. 8 Performance of different algorithms.

6.2 Run time

Different algorithms are implemented; the run time of each algorithm is the most important factor to known which one is much suitable with respect to time to run the application. The run times of each algorithm in seconds described in table 3. Each algorithm is applied on the same features which include the cumulative feature set +event which include different features are *TokN*, *Dir*, *Str*, *POS*, *Inter*, and *Event*.

The C4.5 weka algorithm spends more time to classify the data and extract the relation than other algorithms and the accuracy of the overall F1 measures not perfect very well comparing to other algorithms. The naive bayes weka algorithm needs small time compared to C4.5 weka but the overall F1 value is small than the overall F1 value of C4.5 weka. SVM algorithm is less in time than C4.5 and naïve bayes weka and the F1 measures is greater than these algorithms. This means that SVM is better than C4.5 and Naïve Bayes weka. KNN weka algorithm requires small run time comparing to C4.5, Naïve Bayes, and SVM algorithms. But the accuracy of overall F1 measure is less than SVM and better than C4.5 and Naïve Bayes. The PAUM algorithm considers the faster algorithm than other algorithms and their accuracy F1 measure is better than other algorithms except for SVM algorithm. PAUM algorithm is a faster on small data set than SVM algorithm and there is small difference on the accuracy of the overall F1 measures in between. Figure 9 shows the graph of the run time of each algorithm.

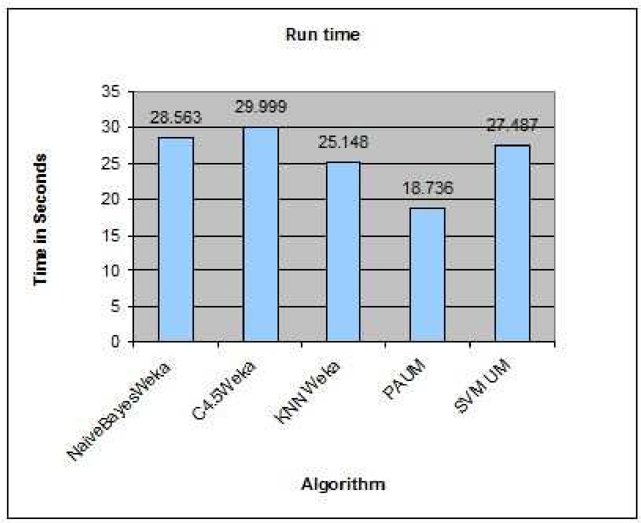

Fig. 9 Run times of different algorithms.

6.3 Uneven margin parameter

SVM with uneven margin is the better algorithm in the accuracy than other algorithm but not the faster one. The SVM algorithm is implemented with different uneven parameters to obtain the value of uneven margin (τ) that improves the performance of the system. Then SVM with this uneven margin value is applied with different features to see the effect of adding new feature to the model and also use different corpus size to known their effects on the performance of the system. Table 4 shows SVM with different uneven parameter values. The standard SVM use the uneven margin value 1, this gives bad results than SVM with uneven margin. When the value of uneven margin parameter decreases the results is improved. Figure 10 describes the graph of using different uneven margin parameters.

When the uneven margin value τ = 0.8 is applied, the performance is improved than τ =1 by around 2.6%. The value of τ is decreased than this level τ =0.8 the value of overall F1 measure decreased. Using τ = 0.6, there is small drop on the value of F1 by 0.1%. SVM with τ = 0.4 effects on the performance, this leads to a drop on the value of overall F1 than τ = 0.6 by around 0.36%. When the value of τ is changed to τ = 0.2, there is drop on the value of overall F1 than τ =0.4 by around 1.46%. This mean that while SVM is implemented with increasing the value of uneven margin τ , the performance of the system is improved until it is reached to the point that the performance is decreased with increasing the value of τ .

Table 4: SVM use different uneven parameter values

	Uneven margin (τ)					
	Metric (%)	1.0	0.8	0.6	0.4	0.2
Overall Relations	P	76.04	73.99	69.58	65.8	61.54
	R	72.16	79.67	85.29	90.78	94.91
	F1	73.7	76.3	76.2	75.84	74.36

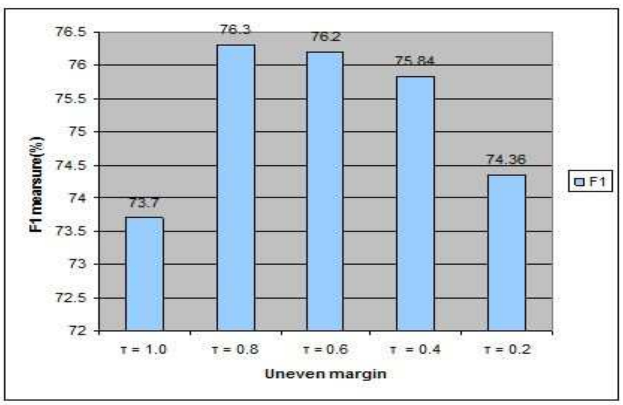

Fig. 10 Performance by different uneven margins.

6.4 Feature selection

The experiments searches for the performance of relation extraction with various feature sets, using the feature sets described in table 2. An additive strategy is used to select the feature. The experiments are divided into two cases, one case of feature sets that do not use syntactic parse information and the other case of feature sets that use syntactic parse information.

6.4.1 Non-syntactic features

Firstly, the experiments used the feature sets that do not use syntactic parse information for relation extraction. Starting with the basic features and then adding new feature set each time to measure the performance of the system. The results are described in table 5.

Starting with Tok6 and Atype features sets, the overall F1 value is 68.29%. Addition of Dir features leads to improve the performance in most metrics, there is improved in the overall F1 value by around 1.35%. Addition of Str features improves the performance in most metrics, there is improved in the overall F1 value by around 0.5%. Addition of the POS features leads to drop the performance in some metrics, overall F1 value drop by around 0.66%. Addition of the Inter features gives more improvements in all metrics, overall F1value increases by around 6.63%. Addition of the Event features gives more improvements in some metrics, overall F1 value increases by around 0.19%.

Generalizing features are used to see their effects on the performance of relation extraction. All Str features, POS features, and TokN features are replaced with their root features, generalized POS features, and generalized TokN features respectively. These results shown in the column Allgen, there is no change in overall F1 value. Notok features are implemented to see if it improves the performance. In this feature TokN features are removed from the full cumulative feature set, corresponding to column +event of table 5. These results are shown in the column Notok, this leads to drop the performance in some metrics, the overall F1value drop by around 0.71%. The graph of using non-syntactic feature sets is shown in figure 11.

6.4.2 Syntactic features

The second part of the feature selection experiments is using features that used syntactic parse information that derived from dependency parse analysis of the texts by using the Stanford parser [43]. The results of +event column in table 5 which corresponding to collection of all non-syntactic feature sets is copied to add in table 6 and then add the Dep features and Syndist features. Addition of the Dep features leads to drop the results that unclear. Addition of the Syndist features leads to a small drop in overall F1 that is unclear. Figure 12 shows the graph of the performance of syntactic feature sets. Addition of the Dep features leads to a drop the performance in some metrics, the overall F1value dropping by around 0.37%. Addition of the Syndist feature set leads to a drop the performance in some metrics, the overall F1 value dropping by 0.38%.

Table 5: Performance by non-syntactic feature sets

Relationship type	*Metric (%)*	*Tok6+ Atype*	*+Dir*	*+Str*	*+POS*	*+Inter*	*+Event*	*Allgen*	*NoTok*
Has_finding	P	5.25	55.16	66.83	45.83	61.66	76.66	76.66	74.76
	R	43.21	48.54	45.69	29.76	42.85	55.85	55.85	58.35
	F1	40.1	44.89	47.06	28.35	43.26	57.32	57.32	58.39
Has_indication	P	64.24	62.34	63.22	62.49	64.56	67.17	67.17	66.04
	R	68.48	69.5	70.16	71.1	69.86	72.85	72.85	71.43
	F1	65.26	64.76	65.37	65.47	66.16	68.89	68.89	67.52
Has_location	P	65.4	65.59	64.98	64.08	78.09	76.68	76.68	76.39
	R	85.5	85.24	87.54	90.3	85.68	85.67	85.67	85.67
	F1	73.38	73.39	73.91	74.4	80.82	80.11	80.11	79.94
Has_target	P	57.04	57.47	58.64	58.18	69.1	68.92	68.92	66.7
	R	67.52	76	76.96	73.68	79.9	76.45	76.45	75.74
	F1	60.08	64.31	65.52	63.77	73.46	71.59	71.59	69.99
Laterality_modifies	P	37.5	45.23	48.16	45	60	60	60	60
	R	37.14	53.57	57.14	47.14	58.57	51.9	51.9	51.9
	F1	36.9	48.47	50.61	45.23	59.23	54.23	54.23	54.23
Negation_modifies	P	70.71	70.71	70.71	70.71	75.71	80	80	80
	R	76.66	76.66	70.83	70.83	71.66	71.66	71.66	71.66
	F1	72.36	72.36	68.93	68.93	71.93	74.66	74.66	74.66
Sub-location_modifies	P	76.54	79.6	79.6	79.6	98.33	100	1.0	1.0
	R	85	93	93	93	93	93	93	93
	F1	78.1	83.02	83.02	83.02	94.64	95.55	95.55	95.55
Overall	P	63.45	63.2	63.45	62.81	73.85	73.99	73.99	73.05
	R	75.39	78.69	79.55	78.97	79.33	79.67	79.67	79.3
	F1	68.29	69.64	70.14	69.48	76.11	76.3	76.3	75.59

6.5 Size of training corpus

Changing the size of training corpus in the experiments is used to examine their effects on relationship extraction. Two subsets with size 20 and 30 documents is selected from 40 documents; referred to them as C20 and C30, respectively.

The collection feature set of *all* non-syntactic feature sets which represent in +event feature set is used in the experiments to show the effects of training corpus size on the performance, these results are shown in table 7.

Firstly, start the experiments with corpus size 20 documents. Increasing the corpus size to 30 documents leads to improve the performance in most metrics; overall F1 value improves by around 2.11%. Using corpus size 40 documents leads to improve the performance in most metrics; overall F1 value improves by around 2.09%. Increasing the size of the training corpus leads to improve the performance of relation extraction system. Figure 13 shows the effects of changing corpus size in the performance.

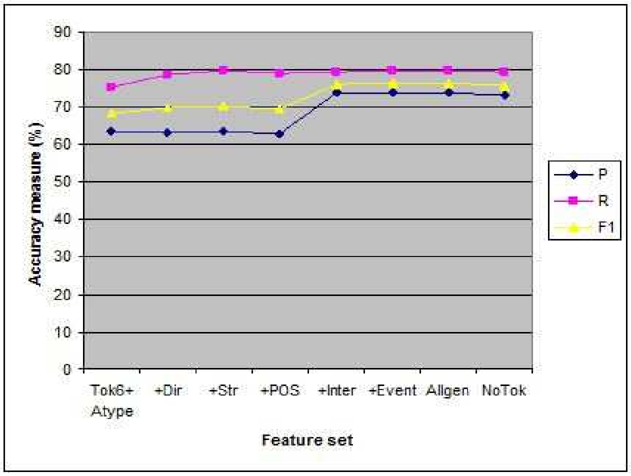

Fig. 11 Graph of non-syntactic feature sets performance.

Table 6: Performance by Syntactic Feature Sets

Relationship type	***Metric (%)***	***+Event***	***+Dep***	***+Syndist***
Has_finding	P	76.66	66.66	45
	R	55.85	44.85	17.02
	F1	57.32	46.49	24.44
Has_indication	P	67.17	66.38	65.37
	R	72.85	71.31	68.53
	F1	68.89	67.77	65.72
Has_location	P	76.68	77.15	79.16
	R	85.67	85.81	85.15
	F1	80.11	80.52	81.12
Has_target	P	68.92	69.23	71.56
	R	76.45	76.84	76.31
	F1	71.59	72.04	73.14
Laterality_modifies	P	60	60	50
	R	51.9	46.9	36.90
	F1	54.23	50.89	40.89
Negation_modifies	P	80	80	80
	R	71.66	71.66	71.66
	F1	74.66	74.66	74.66
Sub-location_modifies	P	100	100	100
	R	93	93	93
	F1	95.55	95.55	95.55
Overall	P	73.99	74	74.93
	R	79.67	78.75	76.98
	F1	76.3	75.93	75.55

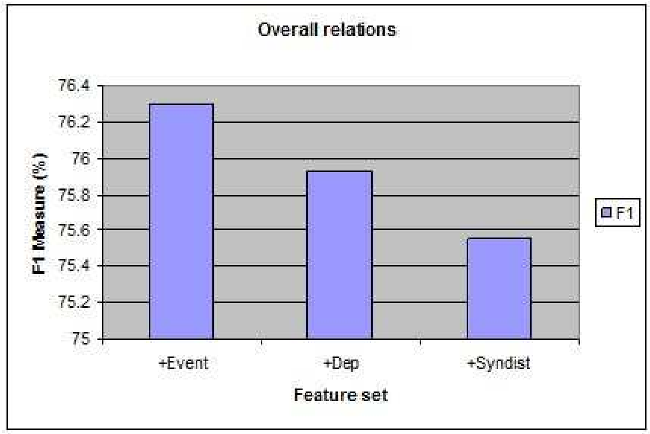

Fig. 12 Graph of syntactic feature sets performance.

7. Conclusion

From the results, the clinical relationships can be extracted from medical text using different supervised machine learning algorithm. SVM with uneven margin is much suitable algorithm which achieves high accuracy, but it takes more time in the run than Perceptron with uneven margin. Perceptron with uneven margin is very fast algorithm than others as well as the accuracy is relatively near to SVM, there is small change in between. SVM with uneven margin is implemented to show the effects of changing the values of *uneven margin* (τ) parameter, adding the feature sets, and changing the size of the training corpus for relationship extraction. Increasing the value of τ leads to improve the performance to reach the value that has high performance where $\tau = 0.8$ after that point the performance dropped. Adding new feature sets like *non-syntactic features* improves the performance. Adding the *syntactic features* leads to small drop in the performance that unclear. Changing the size of training corpus leads to improve the performance. Our future work on relationship extraction in CLEF includes the integration of a noun and a verb chunk tagger into the feature sets.

References

[1] Hahn U& Wermter J, Levels of Natural Language Processing for Text Mining. In: Ananiadou, S. & McNaught, J., ed. Text Mining for Biology and Biomedicine. Artech House: London, 2006.

[2] JISC, Text Mining: Briefing Paper. [Online] Available at: http://www.jisc.ac.uk/media/documents/publications/textminingbp.pdf [Accessed 16 April 2010], 2006.

[3] Rector A, Rogers J, Taweel A, Ingram D, Kalra D, Milan J, Singleton P, Gaizauskas R, Hepple M, Scott D, Power R: CLEF – joining up healthcare with clinical and post-genomic research. Proceedings of UK e-Science All Hands Meeting 2003, Nottingham, UK, 2003:264-267.

Table 7: Performance by corpus size

Relationship type	***Corpus size***			
	Metric (%)	***C20***	***C30***	***C40***
Has_finding	Count	16	23	36
	P	46.66	10	76.66
	R	21.66	3.33	55.85
	F1	29.04	5	57.32
Has_indication	Count	77	125	180
	P	58.69	65.39	67.17
	R	60.22	69.41	72.85
	F1	58.44	65.79	68.89
Has_location	Count	149	239	363
	P	75.05	79.79	76.68
	R	82.4	85.71	85.67
	F1	77.75	82.06	80.11
Has_target	Count	98	145	180
	P	69.4	69.14	68.92
	R	79.52	74.31	76.45
	F1	72.8	70.79	71.59
Laterality_modifies	Count	6	9	15
	P	20	40	0.6
	R	20	30	51.9
	F1	20	33.33	54.23
Negation_modifies	Count	9	11	20
	P	40	0.5	0.8
	R	40	0.45	71.66
	F1	40	46.66	74.66
Sub-location_modifies	Count	11	18	23
	P	60	80	100
	R	60	77.5	93
	F1	60	78.57	95.55
Overall	Count	366	570	817
	P	71.1	73.36	73.99
	R	74.2	75.46	79.67
	F1	72.1	74.21	76.3

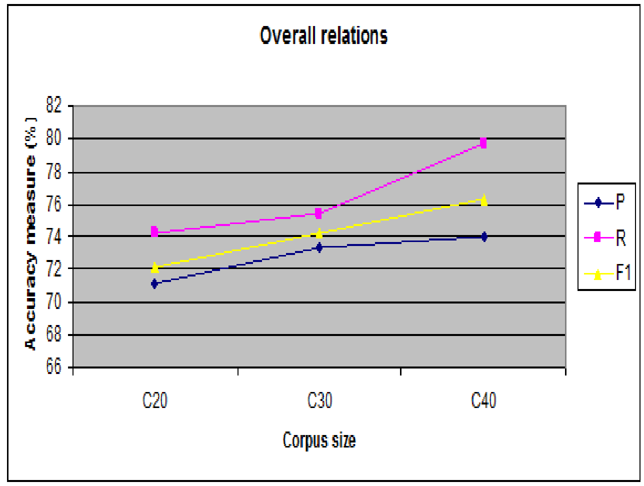

Fig.13. Graph of corpus size performance.

[4] Mitchell T, Machine Learning, McGraw-Hill, 1997.

[5] Roberts A, Gaizauskas R, Hepple M, Guo Y: Combining terminology resources and statistical methods for entity recognition: an evaluation. Proceedings of the Sixth International Conference on Language Resources and Evaluation, LREC 2008, Marrakech, Morocco, 2008.

[6] Turmo J., Ageno A., and Catal`a, N. Adaptive information extraction. ACM Computing Surveys, 2006:38(2).

[7] MUC-5. Proceedings of the 5th Message Understanding Conference, Baltimore, MD, USA. Morgan Kaufmann, San Mateo, CA, 1993.

[8] MUC-6. Proceedings of the 6th Message Understanding Conference, Columbia, MD, USA. Morgan Kaufmann, San Mateo, CA, 1995.

[9] Defense Advanced Research Projects Agency: Proceedings of the Seventh Message Understanding Conference (MUC-7) Fairfax, VA, USA: National Institute of Standards and Technology, 1998.

[10] Chinchor N: Overview of MUC-7/MET-2. Proceedings of the Seventh Message Understanding Conference (MUC-7), Fairfax, VA, USA, 1998.

[11] Doddington G, Mitchell A, Przybocki M, Ramshaw L, Strassel S, Weischedel R: The Automatic Content Extraction (ACE) Program – Tasks, Data, & Evaluation. *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004), Lisbon, Portugal,* 2004:837-840.

[12] Nédellec C: Learning language in logic – genic interaction extraction challenge. *Proceedings of the ICML05 workshop: Learning Language in Logic (LLL05), Bonn, Germany,* 2005:31-37.

[13] BioCreAtIvE II–Protein-Protein Interaction Task [http://biocreative.sourceforge.net/biocreative_2_ppi.html]

[14] Sager N, Lyman M, Bucknall C, Nhan N, Tick L: Natural language processing and the representation of clinical data. *J Am Med Inform Assoc,* 1994:1(2):142-160.

[15] Friedman C, Alderson P, Austin J, Cimino J, Johnson S: A General Natural-language Text Processor for Clinical Radiology. *J Am Med Inform Assoc,* 1994:1(2):161-174.

[16] Lussier Y, Borlawsky T, Rappaport D, Liu Y, Friedman C: PhenoGO: assigning phenotypic context to gene ontology annotations with natural language processing. *Pac Symp Biocomput,* 2006:64-75.

[17] Hahn U, Romacker M, Schulz S: medsynDikate – a natural language system for the extraction of medical information from findings reports. *Int J Med Inform,* 2002:67(1–3):63-74.

[18] Zweigenbaum P, Bachimont B, Bouaud J, Charlet J, Boisvieux JF: A multi-lingual architecture for building a normalised conceptual representation from medical language. *Proc Annu Symp Comput Appl Med Care,* 1995:357-361.

[19] Grover C, Haddow B, Klein E, Matthews M, Nielsen L, Tobin R, Wang X: Adapting a Relation Extraction Pipeline for the BioCreAtIvE II Task. *Proceedings of the BioCreAtIvE II Workshop Madrid, Spain,* 2007.

[20] Bundschus M, Dejori M, Stetter M, Tresp V, Kriegel HP: Extraction of semantic biomedical relations from text using conditional random fields. *BMC Bioinformatics,* 2008:9(207).

[21] Bunescu R, Ge R, Kate R, Marcotte E, Mooney R, Ramani A, Wong Y: Comparative experiments on learning information extractors for proteins and their interactions. *Artif Intell Med,* 2005:33:139-155.

[22] Goadrich M, Oliphant L, Shavlik J: Learning to Extract Genic Interactions Using Gleaner. *Proceedings of the ICML05 workshop: Learning Language in Logic (LLL05), Bonn, Germany,* 2005:62-68.

[23] Giuliano C, Lavelli A, Romano L: Exploiting Shallow Linguistic Information for Relation Extraction from Biomedical Literature. *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL06), Trento, Italy*, 2006:401-408.

[24] Zelenko D, Aone C, Richardella A: Kernel Methods for Relation Extraction. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), Philadelphia, PA, USA,* 2002:71-78.

[25] Zhou G, Su J, Zhang J, Zhang M: Exploring Various Knowledge in Relation Extraction. *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics, Ann Arbor, MI, USA,* 2005:427-434.

[26] Bunescu RC, Mooney RJ: A Shortest Path Dependency Kernel for Relation Extraction. *Proceedings of the Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing,* 2005:724-731.

[27] Roberts A, Gaizauskas R, Hepple M: Extracting Clinical Relationships from Patient Narratives. In Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing Columbus, OH, USA: Association for Computational Linguistics, 2008:10-18.

[28] Roberts A, Gaizauskas R, Hepple M, Guo Y: Mining Clinical Relationships from Patient Narratives. In *Proceedings of the BioNLP 08 ACL Workshop: Themes in biomedical language processing*, 2008: 9(Suppl 11):S3.

[29] Lindberg D, Humphreys B, McCray A: The Unified Medical Language System. Methods Inf Med, 1993: 32(4):281-291.

[30] Roberts A, Gaizauskas R, Hepple M, Demetriou G, Guo Y, Setzer A, Roberts I: Semantic Annotation of Clinical Text: The CLEF Corpus. Proceedings of Building and evaluating resources for biomedical text mining: workshop at LREC 2008, Marrakech, Morocco, 2008.

[31] The University of Chicago medicine, URL http://www.uchospitals.edu/index.shtml.

[32] Mayo Clinic hospital, URL http://www.mayoclinic.com.

[33] Cunningham H, Maynard D, Bontcheva K, Tablan V: GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics, Philadelphia, PA, USA, 2002: 168-175.

[34] URL http://nlp.shef.ac.uk/clef/TheGuidelines/TheGuidelines.html.

[35] Rindflesch T, Fiszman M: The interaction of domain knowledge and linguistic structure in natural language processing: interpreting hypernymic propositions in biomedical text. J Biomed Inform, 2003:36(6):462-477.

[36] K. Ming Leung: Naive Bayesian Classifier. POLYTECHNIC UNIVERSITY Department of Computer Science / Finance and Risk Engineering, 2007.

[37] Quinlan J. R., C4.5: Programs for Machine Learning, Morgan Kaufmann, Los Altos, 1993.

[38] Xindong Wu, et al. "Top 10 algorithms in data mining," Knowl Inf Syst, 2008:14:1–37.

[39] Li Y, Zaragoza H, Herbrich R, Shawe-Taylor J, and Kandola J. The Perceptron Algorithm with Uneven Margins. In *Proceedings of the 9th International Conference on Machine Learning, (ICML-2002)*:pages 379–386.

[40] Li Y, Bontcheva K, Cunningham H: SVM Based Learning System for Information Extraction. In Deterministic and statistical methods in machine learning: first international workshop no. 3635 in Lecture Notes in Computer Science, Springer, 2005:319-339.

[41] Zhou G, Su J, Zhang J, Zhang M: Exploring Various Knowledge in Relation Extraction. Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics, Ann Arbor, MI, USA, 2005:427-434.

[42] Wang T, Li Y, Bontcheva K, Cunningham H, Wang J: Automatic Extraction of Hierarchical Relations from Text. In The Semantic Web: Research and Applications. 3rd European Semantic Web Conference, ESWC 2006 no. 4011 in Lecture Notes in Computer Science, Springer, 2006:215-229.

[43] Klein D, Manning CD: Accurate unlexicalized parsing. In Proceedings of the 41st Annual Meeting on Association for Computational Linguistics Morristown, NJ, USA: Association for Computational Linguistics, 2003:423-430.

Wafaa Tawfik Abdel-moneim is a teaching assistant in information system department in faculty of computers and informatics in zagazig university.

Mohamed Hashem Abdel-Aziz is a professor in information system department in faculty of computers and informatics in ain-shames university.

Mohamed Monier Hassan a professor assistant in information system department in faculty of computers and informatics in zagazig university.