Semantic Networks and Spreading Activation Process for QA improvement on text answers

José Saias[1], **Paulo Quaresma**[1]

[1] Departamento de Informática
Universidade de Évora – Portugal

{jsaias, pq}@di.uevora.pt

Abstract. *Question Answering (QA) systems try to find precise answers to natural language questions. QA extraction result is often an amount of text candidate answers which requires some validation and ranking criteria. This paper presents an automatic answer appreciation technique where extracted candidate answers are represented in a question dedicated associative knowledge base, a semantic network. A spreading activation algorithm looks for semantically related candidate answers, that reinforce each other. The purpose is to enhance the best answers by rising their weight. This article concludes with evaluation details for an experiment with text answers to Portuguese questions.*

1. Introduction

Question Answering (QA) systems receive natural language written questions and look for exact answers. If we ask "*Que animal é o Cocas?*" (in English: *What kind of animal is Cocas?*), a QA system could answer: *peluche* (*plush doll*), *rã* (*true frog*), *sapo* (*toad*), *verde* (*has the green color*), or even *porco* (*pig*) if found as the specie of some pet named *Cocas* (the Portuguese name for *Kermit The Frog*, from *The Muppet Show*). Most open domain QA systems look for answers in Web documents or plain text files [Forner et al. 2008], possibly in static and local collections [Carvalho et al. 2009] [Amaral et al. 2008] [Saias and Quaresma 2008]. Candidate answers are expressions that fit into the expected type of answer for the question category that are related to the question focus, in some document. To handle multiple acceptable answers within a wide list of bogus results is a challenge. Even morphologically different answers might have some semantic connection that can be useful for an informed choice. We propose a method to increase QA systems accuracy by executing an answer appreciation phase over the extraction result that goes beyond the usual QA answer validation.

2. Method

While deciding which possible answer is the best for a question, one considers remembrances of those answer concepts in memory and all known meanings for a term. Inspired by that human process [Duch et al. 2007], our methodology consists on building an associative structure, where all possible answer meanings can be represented along with other semantic information possibly relevant to the question. Within this structure, semantic resemblance can be detected between answers. This can be seen as local consensus or mutual reinforcement, suggesting that those answers are more plausible and thus may be promoted. This method is language and domain independent. Firstly, we create nodes for each factoid answer. Next step looks for the first level of semantic information for those nodes and will connect them to new nodes. Each link from one node to another has a type that represents a semantic relation. Links also have a weight w_{rel}, such that $0 < w_{rel} <= 1$, with the expected confidence degree for that relation. New nodes are obtained with a methodology inspired by previous dictionary based work [Oliveira et al. 2008] and by the SESEMI system [Vanderwende 1995]. An on-line dictionary is used to get a definition for each factoid. A rule based module reads the definition syntactic tree, from the morpho-syntactical analyzer PALAVRAS [Bick 2000], considering nouns, verbs and adjectives whose disposition suggests semantic relations hyponymy, meronymy, synonym and antonym. The definition for the first answer in the example is '**peluche**: *s.m. - Boneco revestido com tecido felpudo*'. Its structure suggests a relation between *peluche* and its definition core noun element: *boneco* (in English: *doll*). The last step is the node expansion. For each existing node, related concepts and semantic relations are imported from associative information resources, such as ontologies or synonym tables. The resulting semantic network is drawn in figure 1. Longer answers with definitions or descriptions begin by identifying the key elements in their text, with a syntax based approach, as described in [Saias 2010]. The full answer node will connect to these key element nodes, which will be expanded through the process mentioned before.

The spreading activation method is inspired in Information Retrieval work from Fabio Crestani [Crestani 1997]. Each answer node has an activation level (AL) equal to his answer weight at QA extraction phase. Remaining nodes start with $AL = 0$. A node can be in active or inactive status. The answer space activation threshold (AT) is 80% of the twentieth best answer weight, or the last if there are less than 20 results. The active state is achieved if $AL >= AT$. When the algorithm is applied to an answer node, only its activation level is considered and propagated to his neighbor nodes as described later. If other nodes become active and have not yet been processed, then the spreading activation process is applied to them. Each relation semantics determines also a propagation coefficient to manage how intense the propagation will be for that relation. These coefficient values can be found at [Saias 2010]. A neighbor node receiving propagation value $PropV$ will update its activation level according to the first equation on the left of figure 2. The starting activation level AL_0 is added to an amount with two components: a

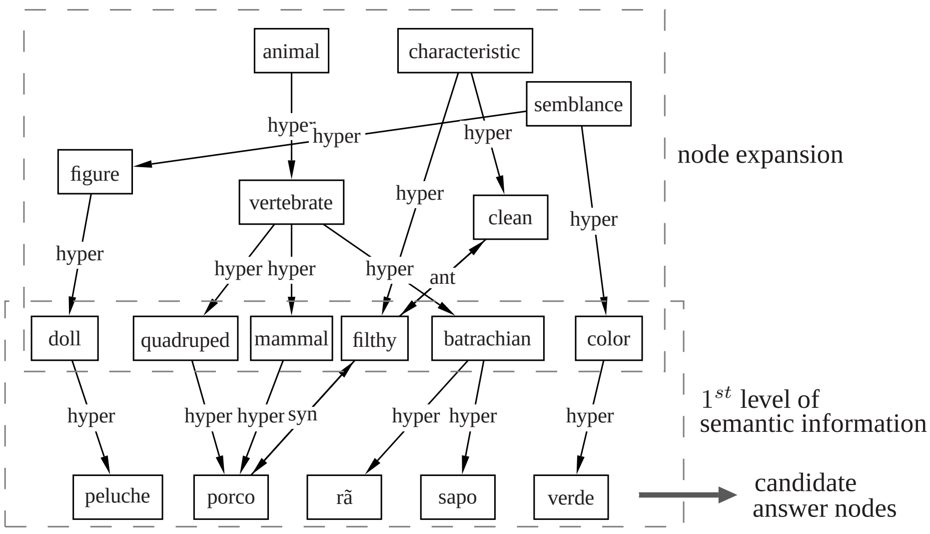

Figure 1. Candidate answers semantic network

$$AL = AL_0 + \sum_{j=1}^{\#props} \left(K \times \frac{RPC_j}{|RPC_j|}\right) + \frac{\sum_{j=1}^{\#props} PropV_j}{\#props}$$

$$PropV = RefAct_{origin} \times W_{rel} \times RPC$$

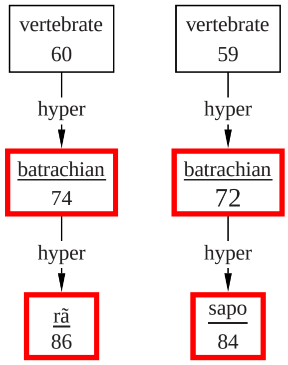

Figure 2. AL propagation formulas and semantic activation spectrum

multiple of an integer value K to reflect the quantity of propagations the target node has received; and $PropV$ average. RPC represents the relation propagation coefficient. The RPC fraction will multiply K per 1 or -1, reflecting that semantic relation's activation reinforcement or inhibition. After some tests, the constant 5 was chosen for K. The value propagated to a neighbor node depends on the relation coefficient and a reference activation level $RefAct_{origin}$. If the propagation source is the answer node doing the spreading process, this reference value is that node activation level. For other cases $RefAct_{origin}$ is the value that the origin node received before as $PropV$. W_{rel} is the relation weight. The semantic activation spectrum for an answer node A is the semantic network segment of nodes that became in active status with A's spreading activation process. The answer space in figure 1 has activation threshold 65. Nodes with underlined text in figure 2 are in active status, after executing the spreading activation process for the two answers shown. This means that the only active node in *sapo*'s spectrum is *batrachian*, shown on the rigth. We compare each semantic activation spectrum and find out that there is one node in common to two answer spectrums. This means that answers *rã* and *sapo* have some semantic affinity. Having such semantic proximity, we can say that these two answers provide a mutual reinforcement. Thus, they can receive an increase of their weight.

Another technique based on the same principle is to apply the spreading activation process to all answer nodes, in order to obtain a combined activation spectrum. The answer nodes connected to these hyperactive nodes can receive a weight increase because they are close to something with semantic relevance. Applying the combined activation spectrum method to the example, we got 78 as the highest activation level on non answer nodes, for *batrachian*. That raises the weight of closer answers *rã* and *sapo*, once more.

3. Experiment

To evaluate the effect of this method we set up an experiment with a collection of *Definition* and *What* questions, written in Portuguese language. After the extraction phase all candidate answers were stored in a database, along with their weight, and then assessed by a human as correct or incorrect. Then we collected some statistic elements for each question: hit on system (first) answer (yes/no); first correct answer on rank (FC); first wrong answer on rank (FW); hit rate in top 5 and in top 10. *Definition* questions had a total of 742 candidate answers, from which 112 (or 15%) were short expressions considered factoids. For category *What*, we had a collection of 316 results, from which 174 (or 55%) were factoid answers. Afterward, we applied the semantic network based answer appreciation method for answer list reranking. Each answer, eventually having an updated weight, was again stored in a database and the statistic report tool was run. Table

Table 1. Assessment with semantic network based techniques

	Definition					*What*				
Technique	*Correct*	FC	FW	Top5	Top10	*Correct*	FC	FW	Top5	Top10
extraction	72.73%	1.32	2.27	47.27%	41.77%	60.00%	5.00	1.80	26.67%	25.11%
SS	72.73%	1.64	3.05	60.00%	50.86%	73.33%	1.87	2.00	30.67%	30.44%
SS-15	86.36%	1.32	3.36	60.91%	49.95%	80.00%	2.00	2.07	30.67%	29.11%
CS	81.82%	2.23	2.86	55.45%	46.31%	26.67%	5.33	1.40	26.67%	25.78%
CS-15	90.91%	1.36	3.23	60.91%	49.04%	60.00%	3.67	1.87	29.33%	27.78%
BOTH	86.36%	1.55	3.27	62.73%	49.95%	–	–	–	–	–

1 shows the average results for each question category, for both spectrum similarity (SS) and combined spectrum (CS) techniques. Looking at SS technique result, for *Definition* type, we see no improvement in system best answer (the first in the rank), keeping the success rate from the extraction phase, 72.73%. The FC indicator is slightly worse, because the average first correct answer position has fallen to 1.64. Indicators Top5 and Top10 show a progress in the concentration of correct answers at the top. In the second line, SS-15 is the same technique but there is an upper bound to the bonus weight increment that an answer can receive. After some tests, the upper bound that allowed better results was 15. With SS-15 the indicators for system correct answer, FC, FW and Top5 are better than without the bonus limit. The result for technique CS shows a bigger gain. Again, the bonus limited variant of the technique works better. The fifth line shows a combination of SS-15 and CS-15 techniques. It seems to be no benefit over a single technique. Answers for *What* category start with a lower hit rate in Top5 and Top10, at extraction phase. Techniques SS and SS-15 improve all indicators. The CS technique has a distinct behavior in each variant. Pure CS technique deteriorates the system success rate from 60% to 26.67%, although the number of correct answers in the first five remains. With CS-15, the system answer success rate keeps equal to extraction phase, at 60%. The remaining indicators denote a small improvement in results.

4. Discussion

This paper presents an automatic answer appreciation technique based on semantic networks and a spreading activation process. The goal is to enhance the results believed to be the most plausible. Our method allows two techniques, spectrum similarity and combined spectrum, described in two variants each: with and without bonus limit. The system answer success rate may increase up to 20% more with these techniques. The CS results were different in each question class. This can be due to the greater homogenity in *Definition* answers. Answers extracted to category *What* are semantically heterogeneous or scattered. This way the combined spectrum is somewhat vague and undefined. On answer space building, the core elements identification is sometimes unaccomplished. As future work we emphasize a further study to determine the propagation coefficient appropriate to each semantic relation, as well as the impact of propagation in the semantic activation of a neighbor node. We use the semantic affinity to promote answers. This works if most answers are correct (as shown by previous answer comparison based methods [Dalmas and Webber 2007]) or at least if there are no big clusters of semantically tied wrong answers. It is unlikely that a QA system produce such negative clusters, so this method seems to be suitable to improve text answer ranking and to favor correct results.

References

Amaral, C., Cassan, A., Figueira, H., Martins, A. F. T., Mendes, A., Mendes, P., Pina, J., and Pinto, C. (2008). Priberam's question answering system in qa@clef 2008. In *CLEF Workshop*, pages 337–344.

Bick, E. (2000). *The Parsing System "Palavras". Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework.* Aarhus University Press.

Carvalho, G., de Matos, D. M., and Rocio, V. (2009). Idsay: Question answering for portuguese. In *Evaluating Systems for Multilingual and Multimodal Information Access*, Lecture Notes in Computer Science, pages 345–352, Aarhus, Denmark. Springer Berlin / Heidelberg. ISBN 978-3-642-04446-5.

Crestani, F. (1997). Application of spreading activation techniques in information retrieval. *Artificial Intelligence Review*, 11:453–482.

Dalmas, T. and Webber, B. L. (2007). Answer comparison in automated question answering. *Journal of Applied Logic*, 5:104–120. Issue 1: Questions and Answers: Theoretical and Applied Perspectives.

Duch, W., Matykiewicz, P., and Pestian, J. (2007). *Towards Understanding of Natural Language: Neurocognitive Inspirations*, volume 4669, chapter II, page 953–962. Springer Lecture Notes in Computer Science.

Forner, P., Peñas, A., Alegria, I., Forascu, C., Moreau, N., Osenova, P., Prokopidis, P., Rocha, P., Sacaleanu, B., Sutcliffe, R., and Sang, E. (2008). Overview of the clef 2008 multilingual question answering track. In Borri, F., Nardi, A., and Peters, C., editors, *Cross Language Evaluation Forum: Working Notes for the CLEF 2008 Workshop*, Aarhus, Denmark.

Oliveira, H. G., Santos, D., Gomes, P., and Seco, N. (2008). Papel: A dictionary-based lexical ontology for portuguese. In *Computational Processing of the Portuguese Language*, volume 5190 of *LNAI*, pages 31–40. Springer. ISBN: 978-3-540-85979-6.

Saias, J. (2010). *Contextualização e Activação Semântica na Selecção de Resultados em Sistemas de Pergunta-Resposta*. PhD thesis, Universidade de Évora.

Saias, J. and Quaresma, P. (2008). The senso question answering system at qa@clef 2008 - working notes. Technical report, Universidade de Évora.

Vanderwende, L. (1995). *The Analysis of Noun Sequences using Semantic Information Extracted from On-Line Dictionaries*. PhD thesis, Georgetown University.