JSCSE
The International Journal of Soft Computing and Software Engineering [JSCSE], Vol. 3, No. 3, Special Issue:
The Proceeding of International Conference on Soft Computing and Software Engineering 2013 [SCSE'13],
San Francisco State University, CA, U.S.A., March 2013
Doi: 10.7321/jscse.v3.n3.37 e-ISSN: 2251-7545

Imitation of the Human Ability of Word Association

Patrick Uhr, André Klahold
Institute of Knowledge Based Systems
University of Siegen
Siegen, Germany
patrick.uhr@uni-siegen.de

Madjid Fathi
BISC-EECS
UC Berkeley, CA
fathi@eecs.berkeley.edu

***Abstract*— The paper in hand introduces a new concept for imitating the human ability of building word association, henceforth called 'CIMAWA'. We have carried out comprehensive case studies to evaluate the ability for imitating human word association. In this context we have used existing studies to compare and examine the performance of our approach. The results have revealed that CIMAWA imitates human word association very accurately and is superior to existing approaches.**

Keywords: Asymmetrical Word Association, Knowledge Discovery from Text, Text-Mining

I. INTRODUCTION

The growth in the number and variety of data collections and the available digital data is ever increasing [1]. An important part of the available data is unstructured text. There is great interest in text-mining techniques [2] to help users gain knowledge [3] from these vast amounts of documents.

Automated keyword extraction and document summarization [4] from a given text are examples to help user select relevant text documents in a more efficient way.

A promising approach to enhance these text-mining methods is to discover the associations between the keywords. Focusing on the human ability of word association we develop a technical implementation based on large text corpora.

To understand association building, preliminary work in the field of human word association has to be done. Hence, an extensive literature study in this area and our own case study 'Human Word Association' have been carried out. On the bases of findings from literature, and the results of the conducted case studies, we have developed the new 'Concept for the Imitation of the Mental Ability of Word Association' (CIMAWA).

We made several tests to compare CIMAWA with other existing approaches. To obtain results that are independent of our own case study, we have chosen the free association test designed by Russell and Meseck [5] as a reference for the association measures.

The subsequent chapters of the paper in hand are organized as follows: Section II deals with the issue of the human word association including literature study and the case study of 'Human Word Association'. Section III introduces CIMAWA approach to imitate human word association. Other existing approaches are tested against CIMAWA and the results are presented in detail. Finally section IV summarizes the main results of this paper, and highlights potentials for future research on applying CIMAWA.

II. HUMAN WORD ASSOCIATION (HWA)

An intuitive way to analyze Human Word Association (HWA) is to conduct surveys. These surveys are common in the fields of psychology and linguistics. The most popular surveys are the free association test (FAT) and the free association norm (FAN). The procedure of these tests is very simple. Stimulus words are presented to the participants, who are asked to utter or write down the first word that comes to mind that is meaningfully related to the presented stimulus [6].

Numerous FATs were conducted in the last decades and raise questions about how the results should be interpreted and used for research purposes. Nelson, McEvoy, and Dennis [7] answer this question with the claim that results of FAT provide us with information about the nature of free association as a knowledge retrieval task. We conclude that results of association experiments can be utilized as a reference mark for HWA.

Nelson, McEvoy, and Schreiber [8] created one of the most famous collections of word associations. Data collection started in 1973 with a total of more than 6,000 participants, 5,019 stimulus words, and nearly 750,000 responses. Using this data collection, Michelbacher, Evert, and Schütze [9] investigated the character of HWA and argue that there are different types of lexical associations including symmetric and asymmetric types. Furthermore, Michelbacher, Evert, and Schütze [9] exemplify symmetric and asymmetric associations with the following stimulus-response pairs derived from [8]. They claim the association of the word-tuple (bad, good) as symmetric, because their elements

The International Journal of Soft Computing and Software Engineering [JSCSE], Vol. 3, No. 3, Special Issue:
The Proceeding of International Conference on Soft Computing and Software Engineering 2013 [SCSE'13],
San Francisco State University, CA, U.S.A., March 2013
Doi: 10.7321/jscse.v3.n3.37 e-ISSN: 2251-7545

primarily relate to each other with about the same strength. In the USF association norm [8], 75% of the participants give 'good' as the response for 'bad', and 76% of the subjects answer with 'bad' to the given 'good' stimulus. Obviously, the participants of the experiment associate both words with nearly the same strength. That indicates a symmetric association between the words. An example for an asymmetric association is the pair (bird, canary), because one element strongly relates to the other but not vice versa. Based on the USF association norm, Michelbacher, Evert, and Schütze [9] discover that 69% of the subjects give 'bird' as a response for 'canary', while only 6% consider 'canary' as a response for 'bird'. The significant difference between the ratios is an indicator for an asymmetric association.

Steyvers, Shiffrin, and Nelson [10] draw the conclusion that "in the norms, the associative strengths […] are often highly asymmetric where the associative strength in one direction is strong while it is weak or zero in the other direction". Possible reasons for the observed ambivalent character of human word association, including the prototype theory [11], are discussed in [12].

Free association experiments and the related literature reveal that there are evidences for asymmetry in HWA. To prove this tendency and to obtain a greater variety of word pairs, where both 'directions' of association are examined, we conducted a case study specially developed for detecting (a)symmetry effects on HWA. Hence, the participants of the 'Human Word Association' case study were instructed to estimate the strength of the relation between a given word 'A' as stimulus and word 'B' as the associated response. They were asked to evaluate –based on a scale from 1 – 'strong association', 2 – 'weak association' and 3 – 'no association' – how strongly 'B' is associated with 'A'. We divided the study into two experimental series: First, a series with 14 words, followed by an 18 word series. In both series, 20 participants were asked to rate their associations. In each series, every possible word tuple was presented to the subjects. The resulting set of 182 (196 minus the 14 tuples were word 'A' and 'B' are identical) word tuples in the first series, and 306 word tuples in the second series were presented to each participant.

TABLE I.
Evaluation results of 'Human Word Association' test series 1

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1		2.00	2.30	1.20	2.70	2.90	1.05	1.30	2.65	2.05	2.00	2.35	1.70	2.40
2	2.25		2.45	2.40	2.30	2.40	2.60	2.25	2.35	2.65	2.75	2.60	2.25	2.85
3	2.05	2.45		1.75	2.75	2.90	2.25	1.80	2.90	2.70	2.70	2.65	2.75	2.15
4	1.15	2.25	1.90		2.50	2.45	1.35	1.30	2.65	2.75	2.70	2.40	2.30	2.60
5	2.70	2.10	2.90	2.65		1.15	2.95	2.60	1.30	1.20	1.40	2.30	1.90	2.70
6	2.70	2.25	2.85	2.65	1.05		2.75	2.50	1.10	1.25	1.55	2.50	2.05	2.75
7	1.15	2.40	2.40	1.60	2.70	2.85		1.75	2.65	2.20	2.40	2.70	2.70	2.10
8	1.20	2.25	1.90	1.20	2.40	2.50	1.80		2.65	2.55	2.75	2.10	2.65	2.15
9	2.65	2.35	2.80	2.65	1.20	1.20	2.75	2.70		1.40	1.40	2.45	2.15	2.85
10	2.00	2.70	2.60	2.80	1.25	1.20	2.30	2.60	1.50		1.40	2.85	2.80	2.55
11	2.10	2.65	2.70	2.70	1.40	1.25	2.35	2.80	1.45	1.20		2.80	2.80	2.75
12	2.40	2.60	2.80	2.10	2.30	2.40	2.65	2.15	2.25	2.80	2.85		2.75	2.85
13	1.70	2.20	2.90	2.45	1.95	2.20	2.70	2.60	2.05	2.75	2.85	2.75		2.65
14	2.30	2.70	2.25	2.35	2.95	2.65	1.95	2.60	2.95	2.70	2.65	3.00	2.60	

Focusing on (a)symmetry aspects, the case study compares the association scores of each word tuple for discovering differences in the rating. The bigger the difference between the association strength 'A' → 'B' to 'B' → 'A', the stronger the evidence for asymmetry between these associations.

Table 1 shows the results of the conducted case study in test series 1. Test series 2 shows similar results and is therefore not presented in detail. Remarkably numbers from 1 to 14 in the first row/column represent the words in the experiment. Numerical values in Table 1 are the average results of the participants association score in the case study. Given stimulus words are presented in the rows and the associated words are listed in the columns. Accordingly, Table 1 presents the average association strength between stimulus $word_1$ and associated $word_2$ at 2.0 (Table I; row 2, column 3). Investigating the opposite direction and answering the question how much the participants associate $word_1$ with $word_2$, the average rating is at 2.25 (Table I; row 3, column 2).

The difference between the association strength in this case is 0.25 (Table II; row 3, column 2) which indicates an asymmetric relation for this word tuple.

TABLE II.
Highlighted differences in 'Human Word Association' test series 1

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1		0.25	0.25	0.05	0.00	0.20	0.10	0.10	0.00	0.05	0.10	0.05	0.00	0.10
2	0.25		0.00	0.15	0.20	0.15	0.20	0.00	0.00	0.05	0.10	0.00	0.05	0.15
3	0.25	0.00		0.15	0.15	0.05	0.15	0.10	0.10	0.10	0.00	0.15	0.15	0.10
4	0.05	0.15	0.15		0.15	0.20	0.25	0.10	0.00	0.05	0.00	0.30	0.15	0.25
5	0.00	0.20	0.15	0.15		0.10	0.25	0.20	0.10	0.05	0.00	0.00	0.05	0.25
6	0.20	0.15	0.05	0.20	0.10		0.10	0.00	0.10	0.05	0.30	0.10	0.15	0.10
7	0.10	0.20	0.15	0.25	0.25	0.10		0.05	0.10	0.10	0.05	0.05	0.00	0.15
8	0.10	0.00	0.10	0.10	0.20	0.00	0.05		0.05	0.05	0.05	0.05	0.05	0.45
9	0.00	0.00	0.10	0.00	0.10	0.10	0.10	0.05		0.10	0.05	0.20	0.10	0.10
10	0.05	0.05	0.10	0.05	0.05	0.05	0.10	0.05	0.10		0.20	0.05	0.05	0.15
11	0.10	0.10	0.00	0.00	0.00	0.30	0.05	0.05	0.05	0.20		0.05	0.05	0.10
12	0.05	0.00	0.15	0.30	0.00	0.10	0.05	0.05	0.20	0.05	0.05		0.00	0.15
13	0.00	0.05	0.15	0.15	0.05	0.15	0.00	0.05	0.10	0.05	0.05	0.00		0.05
14	0.10	0.15	0.10	0.25	0.25	0.10	0.15	0.45	0.10	0.15	0.10	0.15	0.05	

Due to the fact that we are interested solely in the symmetric or asymmetric character of the associations, only absolute values of the difference of the association strength values for each word tuple are significant. Table II presents these absolute differences for experimental series 1.

For an easier visual analysis of the results, we have grouped the differences into 5 intervals and used gray scaling. The strictly symmetric word tuples i.e. no difference in the evaluation score of the associations, ("0.00") remains white. The resulting intervals of differences are [0.01, 0.1]; (0.1, 0.2]; (0.2, 0.3] and (0.3, ∞]. The darker the grayscale, the stronger the tendency for an asymmetric association.

As a result, the case study strengthens the conclusions of the literature study that a reasonable part of HWA has to be characterized as asymmetric or at least cannot be

The International Journal of Soft Computing and Software Engineering [JSCSE], Vol. 3, No. 3, Special Issue:
The Proceeding of International Conference on Soft Computing and Software Engineering 2013 [SCSE'13],
San Francisco State University, CA, U.S.A., March 2013
Doi: 10.7321/jscse.v3.n3.37 e-ISSN: 2251-7545

characterized as strictly symmetric. Therefore the development of an association concept that imitates HWA needs a concept that covers this asymmetric character.

We have consequently developed the 'Concept for the Imitation of the Mental Ability of Word Association' (CIMAWA) to simulate HWA. Section III introduces CIMAWA with explanation of the mathematical background, and shows the performance compared with existing approaches for the imitation of human association building.

III. IMITATION OF HUMAN WORD ASSOCIATIONS

Humans can express the identical meaning in numerous different ways and language is highly redundant [13]. That is why we have developed the idea of using the word associations provided by CIMAWA in order to reduce ambiguity among commonly used terms.

One of the first studies of word association imitation is 'Analog Network of Word Association' [14] for automatic recognition of statistical word association based on co-occurrences of words. Other approaches use conceptual graphs [15], [3] or genetic algorithms [16] for mining associations of semantic relations between words.

The following subsections introduce the concept of CIMAWA and contrast it with existing approaches. Furthermore, we present the tests made to compare CIMAWA with the performance of other well-known autonomous association measurement methods.

To take the whole spectrum into consideration, we have chosen a symmetrical method and one of the rare asymmetrical methods for comparison with the CIMAWA.

As an example for a symmetrical method, we have used Pointwise Mutual Information ('PMI') [17], [18], [19], [20], and as representative for an asymmetrical association measure, we have selected the approach developed by Wettler and Rapp [21] ('WR standard').

A. Association Measures

To render the results of the different methods comparable, the general setup is standardized in the following tests. As a reference for the methods to be compared and for making the results independent from the case studies conducted by the Institute of Knowledge Based Systems and Knowledge Management, the FAT of Russell and Meseck [5] was used. 331 participants took part in the used FAT, and 100 words were evaluated.

Before presenting the results in detail, a brief description of the evaluated association concepts is provided.

PMI as a symmetrical approach is an information-theoretically motivated measure for discovering collocations [22]. A collocation is a significant co-occurrence of words. PMI was originally defined by Fano [17] as mutual information between particular events and adapted to the area of natural language processing by Church and others [18], [19], [20]. The mathematical definition is shown in (1).

$$PMI_{ws}(x,y) = \log_2 \frac{P_{ws}(x,y)}{P(x) * P(y)} \quad (1)$$

In adaptation to our PMI application of two words, 'x' and 'y' are defined as the logarithm of the co-occurrence of both words in a certain window-size "*ws*", divided by the product of frequencies of both words in the corpus. Co-occurrence is defined as a measure to indicate how often two words occur together in the same document in a certain *ws* [17], [23], [24], [25]. The *ws* defines the number of succeeding words that are considered regarding the co-occurrence.

The symmetrical character of that measurement method ends up in the calculated measure for $PMI_{ws}(x,y)$. The co-occurrence of the word pair is divided by the product of the frequencies of both words, so that $PMI_{ws}(x,y)$ is a measure for the association between words 'x' and 'y' in a symmetrical way.

As an example for an asymmetric association concept the approach by Wettler and Rapp [21] ('WR standard') has been chosen. 'WR standard' defines the association between two words 'x' and 'y' as follows:

$$\tilde{A}(x,y) = \frac{H(x\&y)}{H(y)^{\alpha}} \quad (2)$$

This approach is asymmetric, because the measurement of the association strength is calculated by the co-occurrence of a word pair H(x&y) divided by the frequency of the predicted answer H(y) in the corpus. Accordingly the value $\tilde{A}(x,y)$ interprets the association strength between word 'x' and 'y', but not vice versa. Hence the resulting calculated measure is unilateral and describes solely the association x → y. Since the frequency of the word 'y' is the denominator, words with low frequencies in the corpus have a strong influence on the calculated association value. That is a reason for making distinction of cases in (3) by Wettler and Rapp [21].

$$\tilde{A}(x,y) = \begin{cases} \frac{H(x\&y)}{H(y)^{\alpha}} \; for\; H(y) > \beta * Q \\ \frac{H(x\&y)}{(\gamma * Q)} \; for\; H(y) \leq \beta * Q \end{cases} \quad (3)$$

The best results in their tests are observed with $\alpha = 0.68$, $\beta = 0.000005$ and $\gamma = 0.000005$ [26]. For the co-occurrence measure a *ws* of '25' is recommended by the authors. The Q variable is defined as the total number of words within the corpus.

The International Journal of Soft Computing and Software Engineering [JSCSE], Vol. 3, No. 3, Special Issue:
The Proceeding of International Conference on Soft Computing and Software Engineering 2013 [SCSE'13],
San Francisco State University, CA, U.S.A., March 2013
Doi: 10.7321/jscse.v3.n3.37 e-ISSN: 2251-7545

B. *CIMAWA (Concept for the Imitiation of the Mental Ability of Word Association)*

Taking into account the results achieved in section II the basic idea is to develop CIMAWA as a sort of hybrid, covering symmetric and asymmetric aspects of HWA. Conceptual differences between the tested concepts are shown in Fig. 1.

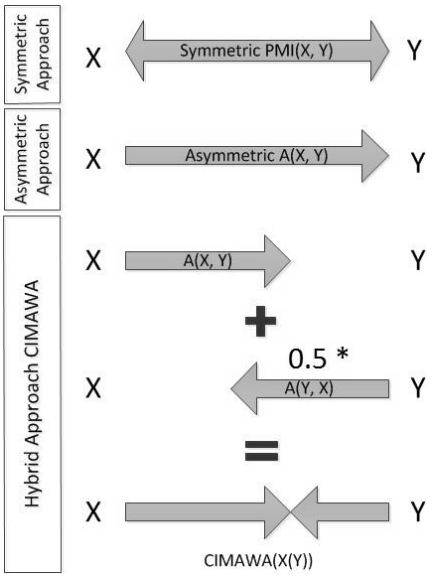

Figure 1. Comparing association measuring concepts

To combine the capability of symmetrical and asymmetrical word association in a comprehensive concept we have carried out a large number of case studies and applied many tests and optimization loops. They are presented in detail in the next section. Consequently, we have derived the following equation from our research:

$$CIMAWA_{ws}\,(x(y)) = \frac{Cooc_{ws}\,(x,y)}{(frequency\,(y))^{\alpha}} + 0.5 * \frac{Cooc_{ws}\,(x,y)}{(frequency\,(x))^{\alpha}} \quad (4),$$

where $CIMAWA_{ws}$ (x(y)) is a measure for indicating strongness of the word 'x' in association with the word 'y', based on a certain *ws*. Aligned to FAT, 'y' is the predicted answer for stimulus 'x' with the word tuple (x (y)) at the highest CIMAWA value.

The equation contains two main parts. First the asymmetric association between the word 'x' and 'y' is calculated. This represents the association between 'x' and 'y' in the direction x → y. In the second part of (4), the inverse direction is considered (y → x). With regard to the results of section 2, which showed that a significant part of HWA is asymmetric, the first summand of (4), which represents the asymmetric part of the word association, gains more weight. Hence, it is included without a damping factor. To make the method hybrid, a second part has been added. That summand represents the symmetric aspect of the association by making use of the association between the predicted answer and the given stimulus. Without the damping factor (0.5) of the second summand the equation would be symmetric, because both summands work like 'looking' from each word in the direction of the other one. In combination to equal summands, we arrive at this symmetric (SYM) equation: (5):

$$SYM\,(x(y)) = \frac{Cooc\,(x,y)}{(frequency\,(y))^{\alpha}} + \frac{Cooc\,(x,y)}{(frequency\,(x))^{\alpha}} \quad (5)$$

which equals to (6) if written with probabilities

$$SYM(x(y)) = \frac{P(x,y)}{P(y)^{\alpha}} + \frac{P(x,y)}{P(x)^{\alpha}}$$
$$\Leftrightarrow\ SYM(x(y)) = (p(x) + p(y)) * \frac{P(x,y)}{P(y)^{\alpha} * P(x)^{\alpha}} \quad (6)$$

Obviously, the second factor represents a slightly modified PMI measure. This PMI resemblance applies entirely to (5), because the first factor has the same value for SYM(x(y)) and SYM(y(x)).
The following section presents the test results of the different association approaches.

C. *Comparative Evaluation*

In what follows the results of the comparative case studies will be presented. The association measures introduced in section III (PMI, Wettler & Rapp) are tested against the CIMAWA approach.
Using the concept of generic algorithms, improvement potentials have been discovered by adjusting the parameter values and *ws*. Partially results could be improved by decreasing *ws* from 25 to 10 and increasing β and γ from 0.000005 to 0.00001087. In our tests these parameter values are defined as 'adjusted' and the original values are named as 'standard'.
The following detailed analysis focuses on two criteria:

1. The prediction of the primary answers, and
2. the average rank, predicted by the different methods, for the primary answer in the FAT.

To render the outcomes of the methods comparable, all results are evaluated by a comparison with the same FAT reference [5].

D. *CIMAWA Case Study 1*

All methods in this case study operate on a corpus which consists of 57,993 editorial texts taken from a German weekly newspaper. The average length of the texts is 1,733 characters.

The International Journal of Soft Computing and Software Engineering [JSCSE], Vol. 3, No. 3, Special Issue:
The Proceeding of International Conference on Soft Computing and Software Engineering 2013 [SCSE'13],
San Francisco State University, CA, U.S.A., March 2013
Doi: 10.7321/jscse.v3.n3.37 e-ISSN: 2251-7545

Starting with an analysis of how many primary answers are discovered by the single methods, Fig. 2 shows the results.

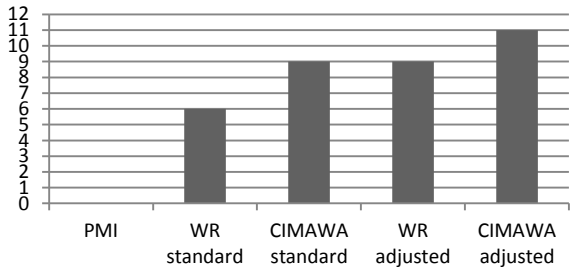

Figure 2. Predicting primary answers

Concerning that criteria, the symmetric PMI approach has achieved the lowest results. The asymmetric approach with standard parameter values ('WR standard') provides six of the top answers when CIMAWA with the same parameter values provides three additional answers. After adjusting the parameter values, the asymmetric method ('WR adjusted') provides nine top answers as well. Nevertheless, the best results are provided by CIMAWA with adjusted parameters. CIMAWA has predicted the correct primary answer eleven times.

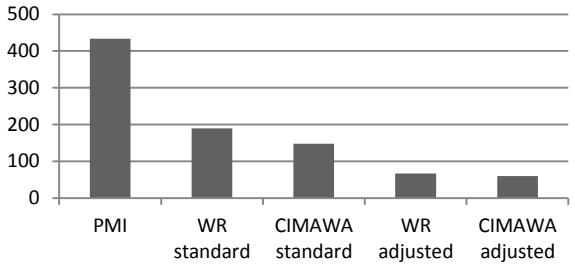

Figure 3. Average rank of the primary answer

The next analysis does not focus on the prediction of completely correct ranks of the answers, but it considers the predicted ranks (Table III; column 3 – 7) of the primary answers given by the subjects (Table III; column 2) in average. Its results are shown in Fig. 3.

Again, the weakest method is the symmetric PMI approach. The primary answer of the FAT is predicted on position 433.5 on average. All other methods have increased the rate of success. Best results are achieved by CIMAWA with adjusted parameters. On average CIMAWA predicted the top answer on position 60.2.

A detailed presentation of the results achieved in case study 1 is presented in Table III. The first column shows the input data for all tested methods and, respectively, the stimulus for the subjects of the FAT. Primary answers of the test persons are given in the second column and the predicted ranks calculated by different association measures are displayed accordingly. All test results are presented where at least one of the tested methods has shown proper results. CIMAWA with adjusted parameter values achieves the best results compared to all other methods in this case study.

The International Journal of Soft Computing and Software Engineering [JSCSE], Vol. 3, No. 3, Special Issue:
The Proceeding of International Conference on Soft Computing and Software Engineering 2013 [SCSE'13],
San Francisco State University, CA, U.S.A., March 2013
Doi: 10.7321/jscse.v3.n3.37 e-ISSN: 2251-7545

TABLE III.

Predicted ranking of the tested methods in case study 1

Stimulus	Primary answer FAT	PMI	WR standard	CIMAWA standard	WR adjusted	CIMAWA Adjusted
butter (butter)	brot (bread)	188	27	19	12	6
rot (red)	grün (green)	117	1	1	1	1
dunkel (dark)	hell (bright)	48	1	1	1	1
musik (music)	ton (tone)	185	3	2	1	1
weich (soft)	hart (hard)	174	419	435	59	65
essen (toeat)	trinken (to drink)	92	1	1	1	1
berg (mountain)	tal (valley)	157	10	7	1	1
haus (house)	hof (yard)	1676	319	302	269	266
obst (fruit)	gemüse (vegetable)	59	1	1	1	1
süß (sweet)	sauer (sour)	77	2	1	1	1
kalt (cold)	warm (warm)	61	2	1	1	2
langsam (slow)	schnell (fast)	663	540	67	242	37
wünschen (to wish)	weihnachten (christmas)	451	324	315	208	237
fluss (river)	wasser (water)	111	195	9	112	113
fenster (window)	glas (glass)	568	128	106	48	44
bürger (citizen)	staat (state)	716	78	43	42	31
sauer (sour)	süß (sweet)	108	5	9	17	9
erde (earth)	himmel (heaven)	112	6	4	2	4
hart (hard)	weich (soft)	276	258	276	30	42
magen (stomach)	darm (gut)	28	1	1	1	1
gelb (yellow)	rot (red)	154	3	1	2	1
brot (bread)	essen (to eat)	529	1171	596	206	175
licht (light)	dunkel (dark)	216	4	4	2	3
schnell (fast)	langsam (slow)	2590	306	270	122	124
kopf (head)	haar (hair)	253	48	65	206	187
bitter (bitter)	süß (sweet)		17	18		
hammer (hammer)	amboss (anvil)	14	1	1	3	1
laut (loud)	leise (quiet)	3557	1679	1842	244	279
ruhig (quiet)	laut (loud)	278	684	445	278	278
salz (salt)	zucker (sugar)	119	3	3	2	2
käse (cheese)	butter (butter)	269	15	16	7	8
spinne (spider)	netz (net)	10	4	3	2	1
ozean (ocean)	meer (sea)	16	7	7	2	3

E. CIMAWA Case Study 2

In comparison to the first case study the corpus is changed and all associations are therefore calculated completely independent from case study 1. For making the outcome comparable to the first study all parameter values and methods are tested again.

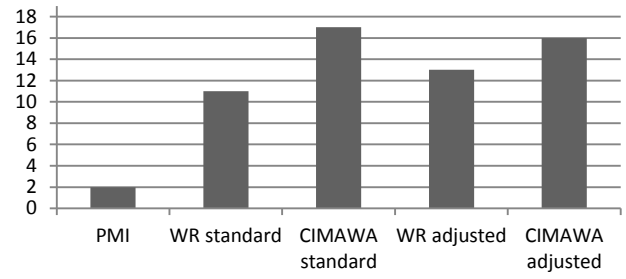

Figure 4. Predicting primary answers

The corpus for this case study is provided by [27] and it consists of approximately 2.8 billion words.

In Fig. 4 and Fig. 5 the results of the second case study are presented concerning the criteria defined before.

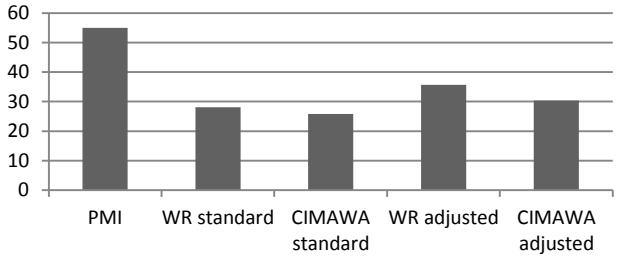

Figure 5. Average rank of the primary answers

Similar to the results of the first case study PMI turned out to be the weakest approach. PMI detects 2 primary answers and ranked the top answers on 55.0253. CIMAWA on standard parameters predicts 17 primary answers in this case study, which turned out to be the best result achieved in all test series. Average ranking was 25.8101 which also proved to be the best result. In comparison, WR standard predicts 11 primary answers and an average ranking of 28.1266.

The International Journal of Soft Computing and Software Engineering [JSCSE], Vol. 3, No. 3, Special Issue:
The Proceeding of International Conference on Soft Computing and Software Engineering 2013 [SCSE'13],
San Francisco State University, CA, U.S.A., March 2013
Doi: 10.7321/jscse.v3.n3.37 e-ISSN: 2251-7545

After adjusting parameter values the results of WR standard improved slightly to 13 primary answers, but results of CIMAWA on the same parameters is still the best measure with 16 primary answers and 30.3291 on average ranking.
Combining the results of both case studies, one can conclude that the best of the evaluated methods for the technical implementation of human word association is the hybrid CIMAWA approach, independent from the chosen parameter values and the used corpus.

IV. Conclusion and Future work

In the framework of the 'Human Word Association' case studies, we have analyzed the character of HWA and, based on these findings, we have developed the new CIMAWA association measure as a technical method to imitate important aspects of HWA. Comparative case studies have shown promising results and proved that CIMAWA can be used as a technical implementation for HWA.
Future research in this area will focus on finding new application areas for the CIMAWA approach for utilizing the discovered potentials of the developed concept.

References

[1] X. J. Ma, W.-X. Wang, Y.-C. Lai, and Z. Zheng, "Information explosion on complex networks and control", *The European Physical Journal B, vol.* 76, *no 1,* pp 179-183, 2010.

[2] J. Dörre, P. Gerstl, and R. Seiffert, "Text Mining: Finding Nuggets in Mountains of Textual Data", *Proceedings International Conference of Knowledge Discovery and Data Mining*, pp. 398-401, 1999.

[3] T. Jiang, A.-H. Tan, and K. Wang, "Mining Generalized Associations of Semantic Relations from Textual Web Content", *IEEE Transactions on Knowledge and Data Engineering, vol. 19, no. 2,* February 2007.

[4] P. Goyal, L. Behera, and T. M. McGinnity, "A Context based Word Indexing Model for Document Summarization," *IEEE Transactions on Knowledge and Data Engineering*, 25 May 2012.

[5] W. A Russell, "The complete German language norms for responses to 100 words from the Kent-Rosanoff word association test.", *Norms of Word Association.* New York: Academic Press, pp 53-94, 1970.

[6] D. L. Nelson, C. L McEvoy, and T. A Schreiber, "The University of South Florida free association, rhyme, and word fragment norms,"*Behavior research methods instruments computers a journal of the Psychonomic Society Inc, Volume: 36, Issue: 3,* pp: 402-407,2004.

[7] D. L. Nelson, C. L McEvoy, and S. Dennis, "What is free association and what does it measure?," *Memory & Cognition*, 28, 887-899, 2000.

[8] D. L. Nelson, C. L McEvoy, and T. A Schreiber, "The University of South Florida word association, rhyme, and word fragment norms," w3.usf.edu/FreeAssociation, 1998.

[9] L. Michelbacher, S. Evert, and H. Schütze, "Asymmetry in Corpus-Derived and Human Word Associations" *Corpus Linguistics and Linguistic Theory*, pre-publication version, 2011.

[10] M. Steyvers, R. M. Shiffrin, and D. L. Nelson, "Word Association Spaces for Predicting Semantic Similarity Effects in Episodic Memory," *In A. Healy (Ed.), Experimental Cognitive Psychology and its Applications: Festschrift in Honor of Lyle Bourne, Walter Kintsch, and Thomas Landauer*, 2004.

[11] E. H. Rosch, "Natural Categories," *Cognitive Psychology*, 1973.

[12] L. Michelbacher, S. Evert, and H. Schütze, "Asymmetric Association Measures", *Proc. of RANLP*, 2007.

[13] C. Lyon, Y. Sato, J. Saunders, and C. L. Nehaniv, "What is Needed for a Robot to Acquire Grammar? Some Underlying Primitive Mechanisms for the Synthesis of Linguistic Ability," *IEEE Transactions on Autonomous Mental Development, Vol. 1, No. 3,* October 2009.

[14] V.E. Giuliano, "Analog Network for Word Association," *IEEE Transactions on Military Electronics*, February 1963.

[15] N. Guarino, C. Masolo, and G.Vetere, "Ontoseek: Content-Based Acess to the Web," *IEEE Intelligent Systems, vol.14, no.3, pp. 70-80*, May/June 1999.

[16] Y. Zhou, J. Du, G. Zeng, and X. Tu, "Constructing tourism association words set based on association rule mining", *Fourth International Conference on Natural Computation*, 2008.

[17] R. M. Fano, "Transmission of information; a statistical theory of communication", MIT Press, New York, 1961.

[18] K.W. Church, W. Gale, P. Hanks, and D. Hindle, "Using Statistics in Lexical Analysis," *Lexical Acquisition: Exploiting On-Line Resources to Build a Lexicon, pp. 115 - 164*, Hillsdale, NJ: Lawrence Erlbaum, 1991.

[19] K. W. Church and P. Hanks, "Word Association Norms, Mutual Information and Lexicography," *ACL27, pp. 76 – 83*, 1989.

[20] D. Hindle, "Noun Classification from Predicate Argument Structures, " *ACL 28, pp. 268 – 275*, 1990.

[21] M. Wettler and R. Rapp, "Computation of Word Associations based on the Co-Occurrences of Words in large Corpora," *Proceedings of the Workshop on Very Large Corpora: Academic and Industrial Perspectives*, Columbus, Ohio, pp. 84-93, 1993.

[22] C. Manning and H. Schütze, "On Foundations of Statistical Natural Language Processing*," The MIT Press* Cambridge, Massachusetts, 1999.

[23] S. Bordag, "A Comparison of Co-occurrence and Similarity Measures as Simulations of Context,"Computational Linguistics and Intelligent Text Processing, *Lecture Notes in Computer Science, Volume 4919/2008*, pp. 52-63, 2008.

[24] I. Dagan, L. Lee, and F. Pereira, "Similarity-Based Models of Word CooccurrenceProbabilities,"*Machine Learning, 34* , Kluwer Academic Publishers, pp 43–69, Boston, 1999.

[25] A. Klahold, "Empfehlungssysteme – Recommender Systems,"*Vieweg + Teubner*, 2009.

[26] R. Rapp, "Die Berechnung von Assoziationen: einkorpuslinguistischerAnsatz," Hildesheim; Zürich; New York: Olms, 1996.

[27] Das Deutsche Referenzkorpus DeReKo, am Institut für Deutsche Sprache, Mannheim, http://www.ids-mannheim.de/projekte/korpora/, last visited 2012-09-28.