Case Representation and Similarity Assessment in a Recommender System to Support Dementia Caregivers in Geriatric and Palliative Care

Sara Nasiri and Madjid Fathi

Department of Electrical Engineering & Computer Science, University of Siegen
Institute of Knowledge Based Systems & Knowledge Management
Hölderlinstr. 3, 57076 Siegen, Germany
sara.nasiri@uni-siegen.de
fathi@informatik.uni-siegen.de

Abstract: In this paper, a case-based reasoner uses International Classification of Functioning, Disability, and Health framework of WHO (ICF codes) and medical expressions to create keyword association profiles. Dementia Case-based Learning Assistant System (DePicT Dementia CLASS) finds significant references and learning materials by utilizing the profile of word association strengths according to the problem description. The purpose of this research is to develop a CBR system for recommending the related references by using the information retrieved from dementia books based on the ICF framework of WHO. Case-based learning assistant system helps users to find their answers in dealing with their problems. It also used their feedback to update cases and for improvement of references. This research proposes a combination of references with the highest value keyword association strengths and collaborative recommendation based on the ranked references by the user's feedback.

Keywords: case-based reasoning, dementia, ICF, palliative care, vocational educational training, adaptation

1. Introduction

A CBR methodology is an approach for the recommendation process in medical applications, and especially in medical assistant systems [1, 2, 3, 4, 5]. CBR applied in various problem-solving domains, and it is appropriate in medicine to integrate the system and for explicit experience, cognitive adequateness, a duality of objective/subjective knowledge, and to extract subjective knowledge [6].
"Dementia encompasses a range of neurological disorders characterized by memory loss and cognitive impairment. In 2015, almost 47 million people worldwide were estimated to be affected by dementia, and the numbers are expected to reach 75 million by 2030, and 131 million by 2050, with the greatest increase expected in low-income and middle-income countries [7]." Therefore, the World Health Organization (WHO) in 2012 and 2015, presented reports that Alzheimer's Disease (AD) and other dementias should be regarded as a global public health priority [7]. International Classification of Functioning, Disability, and Health (ICF) [8] is utilized and developed in different projects for disabilities and health problems. "This was believed necessary due to the

complexity of using a large number of ICF codes needed clinically to classify a person's functioning. Indeed, the maximum number of codes per person can be 34 at the one digit level (eight body functions, eight body structures, nine performance and nine capacity codes). At the second level the number of codes is 362; and, at more detailed levels, these codes total up to 1,424 items [8], [9]." ICF framework is also applied in dementia for matching older adults with dementia and technology. To illustrate the use of the ICF in the clinical management of individuals with dementia [10]. To analyze the communication disorders in Alzheimer [11], and to analyze the prevalence of functional impairments, activity limitations and participation restrictions [12].

Textual case-based reasoning (TCBR) is "a subfield of CBR concerned with research and implementation on case-based reasoners where some or all of the knowledge sources are available in the textual format [13]". It aims to use the textual knowledge with an automated/semi-automated approach for problem-solving. Over the years, there has been significant progress addressing the way of bringing textual knowledge sources into the structured case bases [14], [15], [16]. Bousbahi and Chor proposed a system which is called MOOCs-Rec that recommends appropriate courses of Massive Open Online Courses (MOOCs) from different providers in response to a specific request of the learner [17].

The main objective of DePicT Dementia CLASS is to develop DePicT CLASS concept by enrichment of cases with dementia learning materials (e.g. reference images and textbooks). DePicT CLASS is a case-based learning assistant system to detect and predict disease using image classification and text information [18]. DePicT Dementia CLASS is used and updated by caregivers and domain experts. It enables caregivers and patients' relatives to find their learning materials and references which address the problems that they are looking for. Therefore, searching and finding the appropriate learning materials is significantly requested. Although the increasing prevalence of dementia poses a major challenge for global health at multiple levels [19], CBR is applied in the care of AD patients from 2001 [20].

In this paper, we have another objective to help caregivers and patients' relatives by facilitating the finding of dementia references and learning materials with using a DePicT CLASS's retrieval mechanism based on the word association profile of the request. This research has addressed these questions in the paper; how to create structured case representations from texts and how to evaluate the similarity between textual cases.

This paper is structured as follows. Section 1 presents the introduction and related works. Section 2 explains the DePicT CLASS as a preliminary concept. Section 3 presents the case representation of our system based on the ICF parameters. Section 4, first explains the DePicT Profile Matrix and then makes specification on our case retrieval. Finally, section 5 provides conclusions and future work.

2. DePicT CLASS: Preliminary Concept

This section focused on DePicT CLASS and how graphical and textual information is used as the feature to find appropriate references and learning materials within the CBR case matching, selection and adaptation procedure. DePicT CLASS [18] is a complete cyclic CBR system and integrated process of solving a problem, revising the similar solutions and learning from retained experiences which illustrated in Fig. 1. DePicT

Profile Matrix (1) enriched the knowledge base of DePicT CLASS and in the following, DePicT CLASS procedure system explained that how these features are used as attributes. Afterward, how the requested information of the user request is utilized in the case matching, retrieval, and selection process. Finally, the solutions are adapted based on similar cases for the recommendation which are ranked based on the values of keywords.

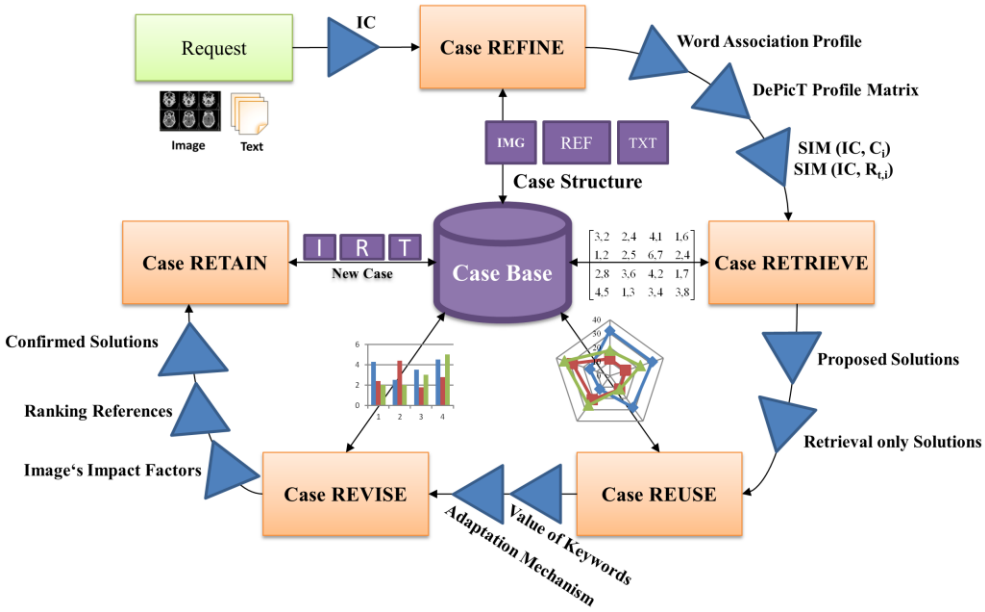

Figure 1: DePicT CLASS Concept [18]

In DePicT CLASS, references and learning materials which are related to the problem and solution are attached to the case as a case description and a case recommendation. Reference images have the word association profile based on the impact factors which are defined by domain experts and tagged to the images. Moreover, DePicT CLASS utilizes collaborative recommendation with tagged keywords, references and learning materials which are ranked by users. The following sections described these principal components of DePicT Dementia CLASS.

3. Case representation based on ICF

Case formation identifies the requested keywords and assigning their values based on DePicT Profile Matrix. This section explained the characterizes a case base and ascertains how incoming cases are refined for retrieving.

Since ICF is inherently a health and health-related classification, it is used as a clinical tool in needs assessment, matched treatments with specific conditions, vocational assessment, rehabilitation and outcome evaluation. It also used as an educational tool in curriculum design and to raise awareness and undertake social action [9]. The structure of ICF is illustrated in Fig. 2 [8].

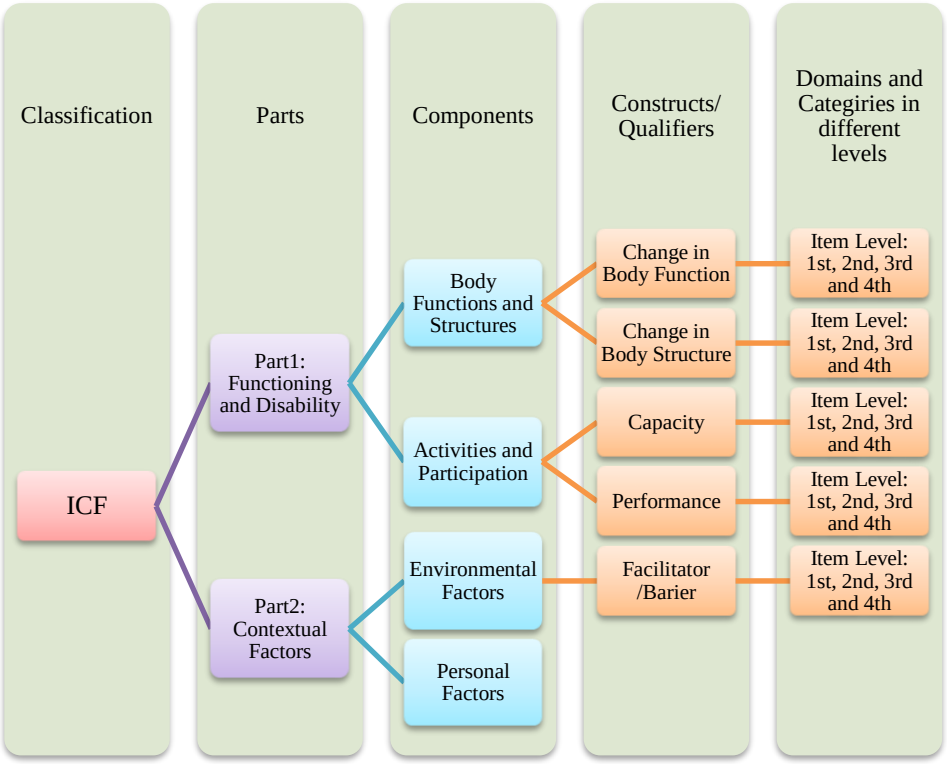

Figure 2: ICF Structure

Scherer et. al. developed ICF codes (111) for dementia with an integrating evidence gathered from preliminary studies that included focus groups of health professionals, a systematic review of the literature, and empirical data collected from patients and caregivers [9]. In this paper, these 111 parameters are utilized as case features. Case representation contains the vectors of ICF word association strengths and ranked common keywords. As shown in Fig 3. case structure includes the identified keywords, problem description, and solution recommendation.

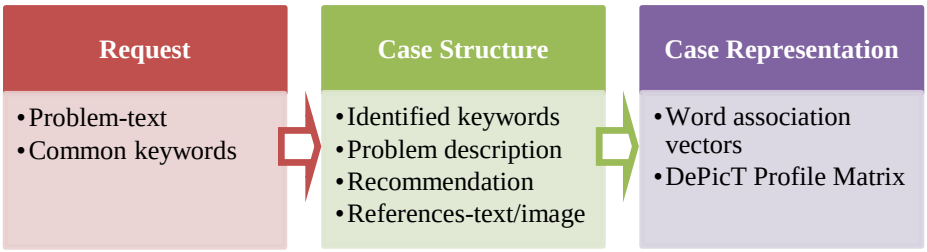

Figure 3: Case Representation

These parameters are searched in dementia, and caregiving books and handbooks e. g. [21, 22, 23, 24, 25, 26, 27] to create the large document as a reference of DePicT Profile Matrix. It is filled based on the ICF words association strength explained in the following section.

4. DePicT Profile Matrix and Case Retrieval

Each case has a word association profile of the main keywords which are defined based on ICF codes and are extracted from case description and case references. The CIMAWA values of the Word Association Strength (WAS) between the case title and case features (identified keywords) are combined in the DePicT Profile Matrix [18].

$$\begin{bmatrix} WAS(1,1) & WAS(1,2) & \dots & WAS(1,n) \\ \vdots & \ddots & & \vdots \\ WAS(j(t,i),1) & WAS(j(t,i),2) & \dots & WAS(j(t,i),n) \end{bmatrix} \quad (1)$$

WAS(j(t,i),i) is the numeric value of CIMAWA [28] between the title phrase of the case i and jth identified keyword of the t^{th} reference from references and learning materials of case i. The case title phrase is a combination of title's keywords as a text string.

$$WAS(x,y) = CIMAWA_{ws}^{\zeta}(x(y)) = \frac{Cooc_{ws}(x,y)}{(frequency(y))^{\alpha}} + \zeta \frac{Cooc_{ws}(x,y)}{(frequency(x))^{\alpha}} \quad (2)$$

The composite character of (2) makes it possible to measure symmetric and asymmetric word associations with a damping factor ζ larger than 0. Co-occurrences ($Cooc_{ws}$) of two words x and y in a defined text window size ws are measured in a large document corpus. These damping factors and window size are changed based on the domain. In this research, to have normalized word association strength (between 0 and 1), best results are achieved with utilizing 2 and 0,5 for α and ζ, respectively and text window size is also ten with five words on the right and five words on the left side of selected keyword. This method considers the identified ICF parameters and its synonyms. The word association strength is also calculated based on 37 surveyed dementia books.

To the list of the ICF identified keywords, the word association strength between Alzheimer and memory loss which is the b144 Memory functions from the ICF second level qualifier is calculated based on the description from Alzheimer's Association [29]:

"**Alzheimer**'s is the most common form of dementia**,** a general term for **memory loss** and other cognitive abilities serious enough to interfere with daily life. **Alzheimer**'s disease accounts for 60 to 80 percent of dementia cases. **Alzheimer**'s is a progressive disease, where dementia symptoms gradually worsen over a number of years. People with **memory loss** or other possible signs of **Alzheimer**'s may find it hard to recognize they have a problem."

According to the equation (1) and (2), the frequency, co-occurrence and WAS of these words are calculated as follows:

$$frequency(Alzheimer) = 4 \quad (3)$$

$$frequency(memory\ loss) = 2 \quad (4)$$

$$Cooc_{10}(Alzheimer, memory\ loss) = 1 \quad (5)$$

$$\mathrm{WAS}(Alzheimer, memory\ loss) = \frac{1}{2^2} + 0{,}5\frac{1}{4^2} = 0{,}28125 \quad (6)$$

For an implementation of this formula, first, the library of pdf box [30] is used. The large text which is created based on ICF parameters for each dementia-related diseases defined as a long string, and it is the string array.
In the second step the frequency of keywords and co-occurrence of them in the ten words (five right and five left) window size is calculated. Therefore, the WAS values are calculated for all keywords in each case as cells of DePicT Profile Matrix(was). Each reference has a word association vector with all relevant keywords of the reference. DePicT CLASS checks the similarity of this vector with the new vector (incoming) which is created with the selected input keywords of a user request. We have also DePicT Profile Matrix(w_i) and DePicT Profile Matrix(w_t) for defining the weights in each case and each reference, respectively.

$$\begin{bmatrix} \mathrm{w11} & \cdots \mathrm{w1j} \cdots & w1k \\ \vdots & \ddots & \vdots \\ \mathrm{wi1} & \cdots \mathrm{wij} \cdots & \mathrm{wnk} \end{bmatrix} \quad (7) \qquad \begin{bmatrix} \mathrm{w11} & \cdots \mathrm{w1j} \cdots & w1k \\ \vdots & \ddots & \vdots \\ \mathrm{wtj} & \cdots \mathrm{wtj} \cdots & \mathrm{wqk} \end{bmatrix} \quad (8)$$

Where $w_{ij} = \frac{f_{ij}}{N}$ is the weight of identified keyword j in the case i and f_{ij} is the frequency of word j in the case i and N are the total number of identified keywords including their frequencies in case i. Moreover, where $w_{tj} = \frac{f_{tj}}{Q}$ is the weight of identified keyword j in the reference t and is expressed as follows:
I) f_{tj} is the frequency of word j in reference t and Q is the total number of common keywords between reference t and IC.
II) Moreover, for the reference image t, f_{tj} is the impact factor of word j in the reference t and Q is the sum of impact factors of all common keywords between a reference image and incoming image.
The similarity measurement for comparison of target case or incoming case (IC) and references in DePicT Dementia CLASS is expressed with the following [18]:

$$\mathrm{SIM}\ (\mathrm{IC}, R_{t,i}) = \sum_{i=1}^{n} \sum_{j=1}^{k} \sum_{t=0}^{q} \frac{w_{tj}.w_{ij}\ (R_{t,i}.\mathrm{IC})}{q} \quad (9)$$

where $R_{t,i}$ is the word association profile vector of t^{th} reference from case i.

$$R_{t,i} = (WAS_{1;0;i};\ :::WAS_{j;t;i}:::;WAS_{r;q;i}) \quad (10)$$

Where $WAS_{j,t,i}$ is the feature value of the word association strength of word j of t^{th} reference in case i. r is the total number of words in the t^{th} reference of case i, and q is the total number of references in case i.
DePicT Dementia CLASS user interface as shown in Fig 4. consists of a query as a free text, list of selected keywords, result and feedback interfaces. The result part contains

the three most similar cases, adapted references, diagram of DePicT Profile of ICF parameters and ranked references.

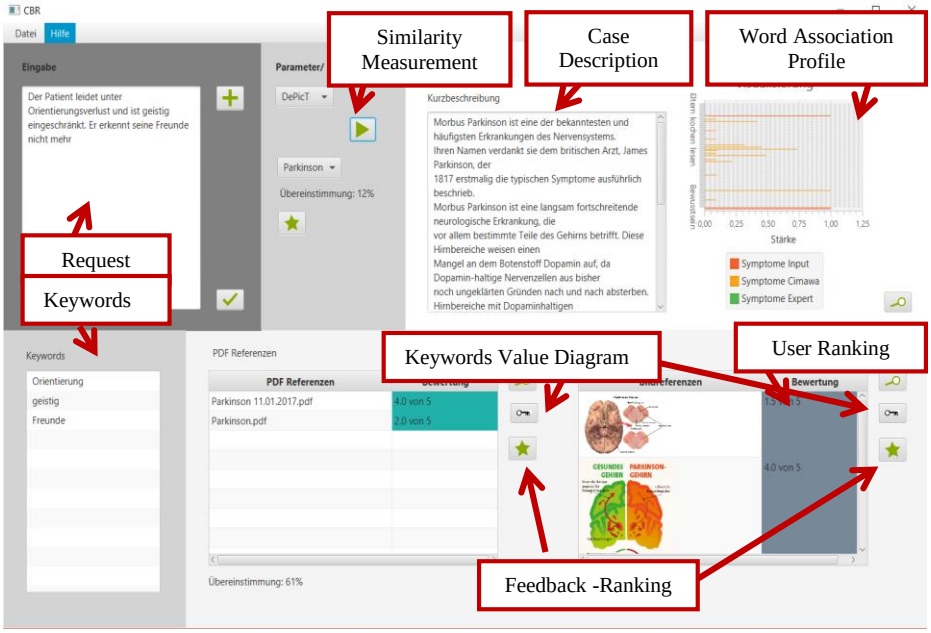

Figure 4: DePicT Dementia CLASS-Case Retrieval:User View

In order to refine the incoming case, IC vector should be created. As an example of the "Requested problem", user request based on the [29] could be as follows:
"It leads to increasingly severe symptoms, including disorientation, mood and behavior changes; unfounded suspicions about family, more serious memory loss and behavior changes; and difficulty speaking, swallowing and walking." Each term is as one element in the list of tokens and the example is represented as follows: [It] [leads][to] [increasingly] [severe] [symptoms] [including] [disorientation] [mood] [and] [behavior] [changes] [deepening] [confusion] [about] [events] [time] [and] [place] [unfounded] [suspicions] [about] [family] [friends] [and] [professional] [caregivers] [more] [serious] [memory loss] [and] [behavior] [changes] [and] [difficulty] [speaking] [swallowing] [and] [walking]. Therefore, based on the ICF identified keywords, common keywords from the requested problem are recognized and IC vector is:

$$IC = [0; \text{disorientation}; 0; \text{memory loss}; 0; \ldots; 0; \text{speaking}; \text{swallowing}; 0; \text{walking}] \quad (11)$$

$$= [0; 1; 0; 1; 0; \ldots; 0; 1; 1; 0; 1] \quad (12)$$

After defining the IC, by utilizing similarity measurement (9), the similarity between IC and each case with its references for these five common keywords is calculated. Similarity degrees of all cases are sorted, and the most similar cases are obtained. However, based on the retrieval only approach, each case which has highest similarity degree selected and its solution should be recommended to the user, in DePicT Dementia CLASS, the highest value references and learning materials of the most

similar cases (the three highest ones) are selected for the recommendation. Therefore, the DePicT CLASS adaptation mechanism has a combination of value comparison based on the requested word association profiles and manual adaptation based on user collaborative recommendation e.g. learner can rank the best references and learning materials based on their understanding and requirements.

5. Conclusion and Future Work

Developing the DePicT Dementia CLASS is the main contribution of this research. It is a case-based system which uses DePicT Profile Matrix of the association strength between title phrase and identified keywords of cases which are dementia related diseases and ICF parameters, respectively. In this analysis, the dementia references and learning materials with high valued keywords in word association profiles from the most similar cases are recommended. The comparison of word association profiles of selected references including image and text recommends high valued associated references to the problem description of an incoming case. Also, the synonyms of word association profiles are created for each case, based on identified keywords as attributes. During the time of using the system, the learning material is ranked and also updated by caregivers and domain experts.The word association strength of keywords is calculated based on the medical document repositories containing thirty-seven dementia books. In future, the other parameters of caregiving e.g. their challenges and task's difficulties will be considered in the features list. Moreover, for evaluation phase, it will be tested with caregivers instead of test problems from Alzheimer and dementia forums and homepages. This research will be extended to the other aspects of this field to supplement domain expert's knowledge on new, complex and unusual cases.

Acknowledgment We would like to thank our students of the WS 2016/17 project group course for their implementation efforts.

References

[1] B. Winblad et. al., "Defeating Alzheimer’s disease and other dementias: a priority for European science and society," *Lancet Neurology Commission*, vol. 15, pp. 455–532, 2016.

[2] K. Bach and K.-D. Althoff, "Developing Case-Based Reasoning Applications Using myCBR 3," in *Case-Based Reasoning Research and Development*, pp. 17-31, 2013.

[3] M. U. Ahmed, S. Begum, E. Olsson, N. Xiong and P. Funk, "Case-Based Reasoning for Medical and Industrial Decision Support Systems," in *Successful Case-Based Reasoning Application*, S. Montani & L.C. Jain (Eds.), SCI 305, Springer-Verlag Berlin Heidelberg, pp. 7-52, 2010.

[4] R. Schmidt, S. Montani, R. Bellazzi, L. Portinale and L. Gierl, "Case-based reasoning for medical knowledge-based systems," *International Journal of Medical Informatics*, vol. 64, no. 2, pp. 355-367, 2001.

[5] S. Nasiri, J. Zenkert, and M. Fathi, "A Medical Case-based Reasoning Approach using Image Classification and Text Information for Recommendation," in *Advances in*

computational intelligence, Rojas et al. (Eds.): IWANN 2015, Part II, Lecture Notes in Computer Science (LNCS 9095), pp. 43-55, 2015.

[6] N. Choudhury and S. A. Begum, "A Survey on Case-based Reasoning in Medicine," *(IJACSA) International Journal of Advanced Computer Science and Applications,* vol. 7, no. 8, pp. 136-144, 2016.

[7] M.-J. Huang, M.-Y. Chen and S.-C. Lee, "Integrating data mining with the case-based reasoning for chronic diseases prognosis and diagnosis," *Expert Systems with Applications,* vol. 32, no. 3, pp. 856-867, 2007.

[8] WHO, "International Classification of Functioning, Disability, and Health (ICF)," 2001.

[9] M. J. Scherer, S. Federici, L. Tiberio, M. Pigliautile, F. Corradi and F. Meloni, "ICF Core Set for Matching Older Adults with Dementia and Technology," *Ageing Int,* no. Springer Science+Business Media, LLC 2010, pp. 414-440, 2012.

[10] T. Hopper, "ICF and Dementia," *Semin Speech Lang,* vol. 28(4), no. PMID: 17935012 DOI: 10.1055/s-2007-986524, pp. 273-282, 2007.

[11] M. B. Badarunisa, D. Sebastian, R. R. Rangasayee and B. Kala, "ICF-Based Analysis of Communication Disorders in Dementia of Alzheimer's Type," *Dement Geriatr Cogn Disord Extra,* vol. 5, pp. 459-469, 2015.

[12] A. Malara, G. Sgrò, F. Ceravolo, G. Curinga, G. F. Renda, F. Spadea and V. Rispoli, "profiles of icf disability in alzheimer and vascular dementia," *Journal of Aging Research and Clinical Practice,* 2012.

[13] R. O. Weber, K. D. Ashley, and S. Brüninghaus, "Textual Case-Based Reasoning," *The Knowledge Engineering Review,* vol. 30, no. DOI: 10.1017/S0269888906000713, pp. 255-260, 2006.

[14] R. Weber, A. Martins, and R. Barcia, "On legal texts and cases," in *In Lenz, M and Ashley, KD (eds) Textual Case-Based Reasoning: Papers from the AAAI-98 Workshop*, Menlo Park, CA: AAAI Press, pp. 40–50, 1998.

[15] K. Ashley, "Applying Textual Case-based Reasoning and Information Extraction in Lessons Learned Systems," in *AAAI Technical Report*, 2000.

[16] K. Gupta and D. Aha, "Towards acquiring case indexing taxonomies from text," in *In Barr, V and Zdravko, M (eds) Proceedings of the Seventeenth Annual Conference of the International Florida Artificial Intelligence Research Society*, Menlo Park, CA: AAAI, 2004.

[17] F. Bousbahi and H. Chor, "Mooc-rec: A case based recommender system for moocs," *Procedia-Social and Behavioral Sciences,* vol. 195, pp. 1813-1822, 2015.

[18] S. Nasiri, J. Zenckert, and M. Fathi, "Improving CBR Adaptation for Recommendation of Associated References in a Knowledge-based Learning Assistant System," *Neurocomputing*, vol. 250, pp. 5-17, 2017.
DOI: http://dx.doi.org/10.1016/j.neucom.2016.10.078

[19] M. Ienca et. al., "Intelligent Assistive Technology for Alzheimer's Disease and Other Dementias: A Systematic Review," *Journal of Alzheimer's Disease,* vol. 56, no. 4, pp. 1301–1340, 2017. DOI: 10.3233/JAD-161037.

[20] S. Marling and P. Whitehouse, "Case-Based Reasoning in the Care of Alzheimer's Disease Patients," *In Lecture Notes in Computer Science*, 2001.

[21] J. W. Ashford and e. al., Handbook of Imaging the Alzheimer Brain, IOS Press BV, 2011.

[22] Y. Mizuno, A. Fisher, and I. Hanin, Mapping the Progress of Alzheimer's and Parkinson's Disease, Springer, 2002.

[23] R. Levine, Defying Dementia: Understanding and Preventing Alzheimer's and Related Disorders, 2006.

[24] S. M. Lobo Prabhu, V. A. Molinari, and J. W. Lomax, Supporting the Caregiver in Dementia, Maryland, United States of America: The Johns Hopkins University Press, 2006.

[25] L. Hardman, Dementia (Diseases and Disorders), Gale, Cengage Learning, 2009.

[26] N. L. Mace, The 36-Hour Day: A Family Guide to Caring for People with Alzheimer Disease, Other Dementias, and Memory Loss in Later Life, THE JOHNS HOPKINS UNIVERSITY PRESS, 2006.

[27] C. Turkington, The Encyclopedia of Alzheimer's Disease, United States of America: ISBN 0-8160-4818-5 (HC), 2003.

[28] P. Uhr, A. Klahold and M. Fathi, "Imitation of the human ability of word association," *International Journal of Soft Computing and Software Engineering(JSCSE)*, vol. 3, no. 3, pp. 248-254, 2013.

[29] Alz.org, "Alzheimer's Association, What Is Alzheimer's?" 2017. http://www.alz.org/alzheimers_disease_what_is_alzheimers.asp

[30] Pdfbox, https://pdfbox.apache.org/, 2017.