Using an Affective Computing Taxonomy Management System to Support Data Management in Personality Traits

Binh Vu
Research Institute for Telecommunication and Cooperation
Dortmund, Germany
bvu@ftk.de

Ryan Donovan
Cork Institute of Technology
Cork, Ireland
brendan.donovan@mycit.ie

Michael Healy
Cork Institute of Technology
Cork, Ireland
michael.healy2@mycit.ie

Paul Mc Kevitt
Ulster University
Derry, Northern Ireland
p.mckevitt@ulster.ac.uk

Paul Walsh
Nsilico Life Science Ltd.
Cork, Ireland
paul.walsh@nsilico.com

Felix Engel
Research Institute for Telecommunication and Cooperation
Dortmund, Germany
fengel@ftk.de

Michael Fuchs
Wilhelm Büchner University of Applied Sciences
Pfungstadt, Germany
michael.fuchs@wb-fernstudium.de

Matthias Hemmje
Telecommunication and Cooperation
Dortmund, Germany
mhemmje@ftk.de

***Abstract*— Affective Computing is a rather new and multidisciplinary research field that seeks sophisticated automation in emotion detection for later analysis. However, the automated emotion detection and analysis require as well comprehensive data management support, e.g. to keep control of data produced, and to enable its efficient reuse through classification with established terminology. This paper contributes to data management aspects in Affective Computing and to automation support in emotion classification on the basis of a personal traits analysis. Hence, we describe the implementation of a taxonomy management system, derived from requirements of a case study that investigates the relationship between personality and emotions in Affective Computing. The study makes use of machine learning software developed by SenseCare, an EU-funded R&D project that applies Affective Computing to enhance and advance future healthcare processes and systems.**

Keywords—Affective Computing, machine learning, emotion detection, taxonomy management system, emotion classification

I. Introduction and Motivation

Affective Computing (AC) is an emerging inter-disciplinary field developing technology that attempts to detect, analyse, and process important psychological aspects such as emotions, feelings, or behaviours with the goal of improving human-computer interaction [1]. *Sensor Enabled Affective Computing for Enhancing Medical Care* (SenseCare) is a 48-month project funded by the European Union, that aims to apply AC to enhance and advance future healthcare processes and systems, especially in providing assistance to people with dementia, medical professionals, and care givers [2]. By gathering activity and related sensor data to infer the emotional state of the patient as a knowledge stream of emotional signals, SenseCare can provide a basis for enhanced care and can alert medics, professional care taking staff, and care taking family members to situations where intervention is required [3] [4].

One of the systems developed in SenseCare is a machine-learning-based emotion detection platform [5], which is used to provide an early insight into the emotional state of an observed person. SenseCare can work from a live video stream or a pre-recorded video, which enables an analysis to be completed on the fly or at a later stage. For automated emotion analysis, we have created a ML model using support vector machine technology called Libsvm[1] and training data from the Cohn-Kanade[2] and Multimedia Understanding Group[3] datasets. To use the generated ML model we developed an application that uses a webcam or video file to feed image frames into the application and it outputs an emotional classification. The outcome of this analysis, i.e. the emotion classification, has to be indexed or classified properly to enable its efficient reuse, and searched and retrieved in a fast, easy, and accurate manner. In psychology, e.g. a taxonomy would permit researchers to study specified domains of personality characteristics, rather than examining separately the thousands of particular attributes that make human beings individual and unique [6]. In the case of classification, a taxonomy can be a controlled vocabulary of emotions, which is a hierarchy of agreed-on terms that later will be used for cataloguing scientific content related to emotions [7]. Whilst several well-known taxonomy management systems exist today, they are not synchronized with SenseCare platforms and user data. Furthermore, these systems are missing a voting and well-equipped version control mechanism to support the collaborative development of psychologically oriented emotion taxonomies along with their evolution processes. Hence, we have enhanced in SenseCare a taxonomy management system based on a prototype that has been deployed in several R&D projects [7] [8] [9]. The application of the developed taxonomy management system to SenseCare is described here. To derive requirements for the taxonomy manager and for evaluation purposes we have conducted a personal traits study, which applied detection of emotions to predict relatively enduring psychological attributes of participants, such as their personality traits. Personality traits are typical patterns of emotion, cognition,

[1] https://www.csie.ntu.edu.tw/~cjlin/libsvm/
[2] http://www.consortium.ri.cmu.edu/ckagree/
[3] https://mug.ee.auth.gr/fed

978-1-7281-1867-3/19/$31.00 ©2019 IEEE

behaviour, and motivation across time [10]. In other words, personality traits reflect how people generally tend to perceive, feel, and act in relation to their environments over long stretches of time. As a result, attempting to understand and or/improve the overall psychological well-being of patients without understanding individual differences means that researchers and health practitioners are restricted to a blanket treatment approach that may only be optimal for a certain group of people. For example, there is evidence to show that a basic psychotherapeutic treatment for Major Depression Disorder (MDD) has differential affects based on the person's personality traits [11]. In a different context, students have been shown to respond better to certain types of curricula and lesson plans, when it is more in-line with their general characteristics [12]. Hence, the ability to provide intelligent and holistic assistance to healthcare practitioners, caregivers, and patients should be improved by understanding personality. The content of this paper therefore is twofold. On the one side, it documents our development and assessment of a taxonomy management system and on the other hand, it provides insights into a study about personality traits. Hence, the paper is organized as follows: section II gives an overview of the state of the art in taxonomy management. Section III introduces the case study on personal traits including the modelling and implementation of the supporting Taxonomy Manager. Section IV discusses the usage of the emotion classification system in SenseCare. Finally, a conclusion is reached in section V.

II. State of the Art

Bruno and Richmond (2003) state that "a taxonomy is a hierarchical classification of headings constructed using the principles of classification, and a thesaurus supplies the commentary and links to navigate the taxonomy" [13]. Hunink et al. (2010) on the other hand defined taxonomy as a structure allowing the classification of entities according to internal criteria, properties, and relationships. Nevertheless, their colleagues do not want to call taxonomy an ontology or simple ontology because it is used for complex structures that have support for automatic reasoning [14]. Daconta et al. (2003) view taxonomy as "a semantic hierarchy in which information entities are related by either the subclassification-of relation or the subclass-of relation. The former is semantically weaker than the latter, so we make a distinction between semantically weaker and semantically stronger taxonomies". Daconta et al. believe that thesauri and ontologies are also taxonomies. They just have stronger semantic relations [15]. One example of a taxonomy used in personality analysis is the Five Factor Model (FFM) [6]. This taxonomy has been developed and extensively used in research by psychologists in the last 30 years. The value the Big Five taxonomy brought to psychologists was a shared framework of assessing the volume of relevant components in personality. The Big Five are made up of the following broad traits:

- *Openness to Experience.* People who score high on this trait tend to be more intellectually curious, creative, and artistically orientated than those lower on this trait.
- *Conscientiousness.* People who score high on this trait tend to be more disciplined, rule-following, and industrious than those lower on this trait.
- *Extroversion*. People who score high on this trait tend to be more sociable, excitable, and assertive than those lower on this trait.
- *Agreeableness*. People who score high on this trait tend to be more compassionate, empathetic, and co-operative than those lower on this trait.
- *Neuroticism*. People who score high on this trait tend to be more emotional volatile, uncertain, and sensitive than those lower on this trait.

Whilst some researchers argue that these classification models are not an optimal domain model of assessing personality, it is a practical and valid classification domain model, as scores on each trait significantly predict important life outcomes, such as: likelihood of experiencing a mental illness, political preferences and ideological preferences, number of romantic partners in life, likelihood of long-term career success, long-term wellbeing, amongst many others [16] [17] [18]. Nevertheless, whilst there has been considerable research investigating the FFM, in recent years, updated models have been proposed that have shown promise of providing higher resolution results. One example of this is Cybernetic Big Five theory, which breaks down each of the five factors into two subcomponents, which are measured with the Big Five Aspects Scale questionnaire [19]. The sub-traits in this model are: *Openness* and *Intellect (Openness to Experience), Industriousness* and *Orderliness (Conscientiousness), Assertiveness* and *Enthusiasm (Extroversion), Agreeableness (Compassion* and *Politeness)*, and *Neuroticism (Withdrawal* and *Anxiety)*. This updated model has shown its ability to clarify previously confusing results in the psychological literature. For example, there had been mixed correlations between the relationship between *Agreeableness* and political preferences. However, when using this updated model, researchers were able to find a positive correlation between left-wing beliefs and the *Compassion* sub-trait of *Agreeableness*, and a positive correlation between right-wing beliefs and the *Politeness* sub-trait of *Agreeableness* [20]. Hence, the ability to detect these sub-traits traits could have important benefits in treating and caring for those who are unable to articulate their own idiosyncrasies through speech or writing. Taxonomy management is the process involved in creating, applying, and maintaining a taxonomy. An individual or organization must first determine a suitable structure for the data it has accumulated or will accumulate in order to create a new taxonomy. Once a new taxonomy is created, it will need to be updated regularly to remain relevant and useful as new information is incorporated and as changes occur in terminology, technology, and markets. In the past, knowledge workers are responsible for creating and maintaining taxonomies. As the workload for knowledge workers increases, many organizations switch to taxonomy management systems. This helps to reduce time expenditure and improve consistency of their information management and classification processes [21].

Taxonomy management systems usually work with content management and enterprise search systems. These three systems help organizations manage their content and knowledge. Taxonomy management systems can be software designed specifically for managing taxonomies or software that have that has other purposes applied to creating and managing taxonomies. Examples of software that were not designed to manage taxonomies but are frequently used for taxonomy management are spreadsheet software, e.g., Excel, and generic database management software, e.g., Microsoft Access. Numerous taxonomies are still created and managed by them. There are various reasons for that, but mostly because specialised taxonomy management systems are costly and complex. Users need to be trained in order to use them. With Excel or Access, most users already use them in the office, so initial time investment is low. Nevertheless, there are also some drawbacks. For example, Excel does not know if the user puts a term in the wrong column. If the taxonomy has thousands of nodes, the user will have difficulty to zoom in and out the taxonomy. It is also inconvenient to move a node with its children to another position because the user has to rearrange terms in the correct column afterwards. PoolParty Taxonomy and Thesaurus Manager is a web-based application in the software package “PoolParty Semantic Suite”. This tool enables the user to build and maintain information architecture. It differs from other market solutions for taxonomy management having powerful capabilities such as corpus analysis. PoolParty enables users to start their work with minimum training. Subject matter experts can model their fields of expertise without IT support. Applying W3C standards, such as SKOS and SKOS-WL, it ensures interoperability of the user’s thesaurus with other taxonomies, knowledge graphs, and APIs. Finally, PoolParty lets users import existing taxonomies and thesauri from Excel or XML. Users can also export them to different standard formats [22]. WebProtégé [23] is a web-based lightweight ontology editor. The goal in developing WebProtégé was not to offer yet another ontology editor, but rather to fill a significant gap in the landscape of ontology tools. The developers wanted to provide an ontology tool that a large spectrum of users, ranging from ontology experts to domain experts, could use. Thus, the ability to customize the user interface for users with different levels of expertise was utmost in its design. Furthermore, WebProtégé provides extensive collaboration support, including change tracking, contextualized threaded discussions, watches and notifications, an extensible access policy mechanism, and generation of statistics on the ontology-development process [24]. Of existing taxonomy management systems on the market, WebProtégé seems to be the best solution. Not only does it support many important features and standards, but it is also an open-source project that enables developers to add new features on top of the current implementation. Nevertheless, one major problem that prevents adopting standalone taxonomy management systems is that they do not have synchronization with business application systems and actual user data. The taxonomy management system must be embedded in a range of other applications in the same platform to support a bigger data governance picture. Hence, an integrated taxonomy manager, that supports important features and standards, and is fully integrated with the current SenseCare content and knowledge management platform, which is implemented by the Content and Knowledge Management Ecosystem Portal (KM-EP) [7], is needed.

III. The Personality Traits Study

In SenseCare, we have conducted a study on personality traits and basic emotions, assessed by both subjective-emotional self-ratings and the automated classification of emotions from initial emotion analysis [1]. The study’s procedure was as follows: First, participants filled out the Big Five Aspect Scale. Second, participants then watched

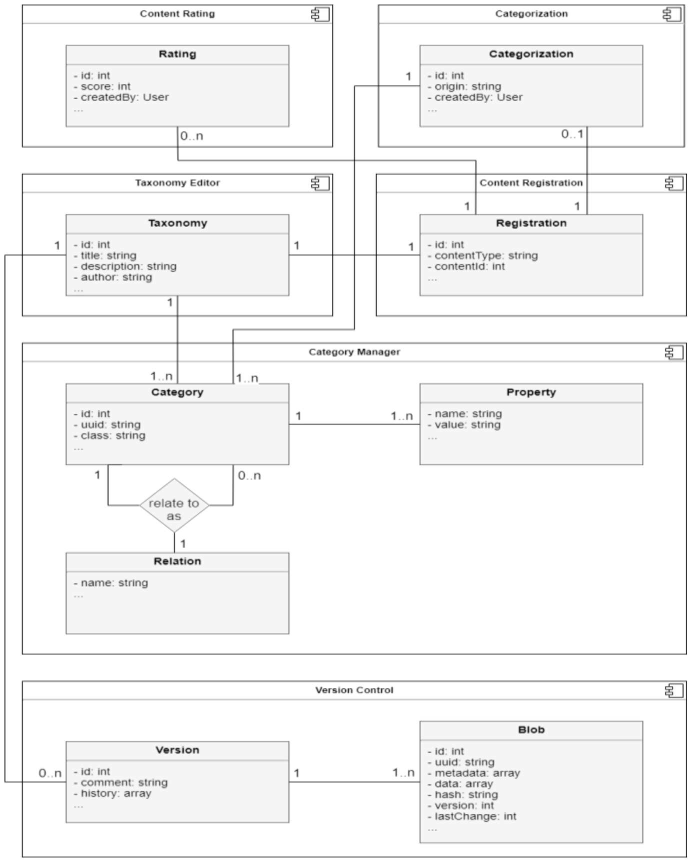

Fig. 1. Conceptual model of Taxonomy Management System

12 video clips taken from movies and TV series, which were designed to cause an emotional response. There were 2 videos per Basic Emotion (i.e. *Disgust*, *Anxiety/Fear*, *Anger*, *Surprise*, *Sadness*, and *Joy*). The video clips lasted between 1 and 5 minutes. Third, after each video clip, participants were asked to rate on a Likert-scale from 1-5 how much of each basic emotion they felt whilst watching the video (1 being “not at all” and 5 being “a great deal”). Participants were debriefed about the nature of the study. Overall, it lasted for an hour and 30 participants took part.

A. Relationship Between Personality Traits and Subjective Emotional Feedback

A Pearson’s correlation was conducted on each personality trait variable with each distinct emotion self-report rating. Pearson correlations measure the relationship between two variables and output the strength of the relationship with the effect size, r. This relationship can range from -1 to +1. A score of -1 represents a perfect negative correlation between two variables. If variables A and B have a negative correlation of -1, every increase of A is matched with an equivalent decrease in B, and vice versa. The opposite is true for a correlation of +1. A correlation of 0 means no relationship exists between the two variables. For social sciences, an effect size of 0.10 is considered small, 0.30 medium, and 0.50 a large effect size. Table 1 shows a correlation matrix between each self-reported emotion and each personality variable. Over half of the relationships have a small-to-medium effect size (59/105). This is in line with previous research in personality and other psychologically important variables.

Table 1. Pearson R Correlation of Big Five Aspect Scale Personality Traits and Self-Reported Emotions

Personality Trait	*Openness to Exp.*	*Openness*	*Intellect*	*Conscientiousness*	*Orderliness*	*Industriousness*	*Extraversion*	*Assertiveness*	*Enthusiasm*	*Agreeableness*	*Politeness*	*Compassion*	*Neuroticism*	*Volatility*	*Withdrawal*
Fear	-0.28	-0.22	-0.26	-0.28	-0.22	-0.26	-0.28	-0.22	-0.26	-0.18	0.01	-0.26	-0.28	-0.22	-0.26
Anger	-0.13	-0.08	-0.14	-0.13	-0.08	-0.14	-0.13	-0.08	-0.14	-0.18	-0.09	-0.18	-0.13	-0.08	-0.14
Joy	0.00	-0.01	0.02	0.00	-0.01	0.02	0.00	-0.01	0.02	-0.38	-0.24	-0.34	0.00	-0.01	0.02
Anxiety	-0.19	-0.10	-0.22	-0.19	-0.10	-0.22	-0.19	-0.10	-0.22	0.00	0.17	-0.14	-0.19	-0.10	-0.22
Sadness	-0.11	-0.01	-0.20	-0.11	-0.01	-0.20	-0.11	-0.01	-0.20	-0.04	0.01	-0.07	-0.11	-0.01	-0.20
Surprise	-0.10	-0.10	-0.06	-0.10	-0.10	-0.06	-0.10	-0.10	-0.06	-0.04	-0.09	0.02	-0.10	-0.10	-0.06
Disgust	-0.15	-0.17	-0.07	-0.15	-0.17	-0.07	-0.15	-0.17	-0.07	-0.06	0.00	-0.08	-0.15	-0.17	-0.07

The vast majority of these relationship are negatively correlated (88/105), meaning that increases in scores of each trait tended to decrease the reported amount of each emotion. The strongest relationship between two variables, was between the personality trait *Agreeableness* and the emotion *Joy*. This seemed to particularly stem from the sub-trait *Compassion*, which also had a moderate negative correlation with *Joy*. In relation to the hypotheses, *Extraversion* and its sub-traits *Assertiveness* and *Enthusiasm* showed no significant or strong relationship with *Joy*; *Neuroticism* was negatively related to the experience of each negative emotion, *Anger, Fear, Anxiety*, *Sadness*, *Disgust*; *Agreeableness*, both *Compassion* and *Politeness* was negatively associated with *Anger*; *Conscientiousness*, both *Industriousness* and *Orderliness*, was negatively associated with *Disgust*; *Openness* to *Experience* had a weak-to-moderate negative correlation with *Anger*, *Fear*, *Sadness*, *Surprise*, *Anxiety*, and *Disgust*, suggesting it is associated with a wide range of emotional experiences. However, it had no relationship with *Joy*.

B. *Relationship between Personality Traits and Automated Emotional Expressions*

Table 2 shows the Pearson R effect size for the correlations between SenseCare ML Emotion Classifications and SR Personality Traits. 10 of the 90 correlations were above r = .20, whilst over half of the correlations were above .10, meaning the vast majority were weak-to-moderate in strength. The highest correlation was between *Agreeableness* and *ML Sadness*, with a moderate negative relationship, r = -0.34. This means that the more agreeable a person scored, the less likely they were to be classified as experiencing sadness during the experiment. Some of the most interesting findings were, *ML Joy* was positively correlated with *Extroversion*, both *Assertiveness* and *Enthusiasm*; *ML Anger* was positively correlated with *Agreeableness*, and both *Compassion* and *Politeness*, in contrast to *SR Anger* which was negatively correlated; *ML Disgust* was negatively associated with *Conscientiousness*, and both *Industriousness* and *Orderliness*, which was similar to *SR Disgust*; *ML Fear/Anxiety* was weakly positively correlated with *Neuroticism*, both *Volatility* and *Withdrawal*; Every ML emotion bar *Fear/Anxiety* had a > .10 (both plus and negative) relationship with *Openness* to *Experience*, suggesting a breadth of emotional experience. However, the strength of each relationship was not matched in the sub-trait aspects of *Openness* to *Experience*, *Intellect* and *Openness*.

C. *Relationship between Subjective Emotions and Facial Expressions*

Table 3 shows the Pearson R correlations between the self-reported emotions and the SenseCare ML detected emotions. As is evident from the table, there are a wide number of discrepancies between the two measures. For example, self-reported *Joy* is positively correlated with *ML Anger*. *ML Joy* is positively correlated with *Fear*, *Anger*, *Anxiety*, and *Disgust*. *SR Disgust* is negatively associated with *ML Disgust*, albeit the size of the correlation is weak. *Fear* is weakly positively correlated with *ML Fear/Anxiety*. *SR Anger* is negatively and weakly correlated with *ML Anger*. There is a weak and positive relationship between *SR* and *ML Joy*. There is a weak-to-moderate positive relationship between *SR Anxiety* and *ML Anxiety/Fear*. There is a weak positive relationship between *SR* and *ML Sadness*. There is a moderate-to-strong positive relationship between *SR* and *ML Surprise*. Overall, there are both consistent and inconsistent results between the two emotions measures.

D. *Taxonomy Management System*

During the study and its outcome, we have derived various requirements that should be supported by a Taxonomy Management System. First, we have investigated the data produced during the study. Secondly, we have derived requirements towards the Taxonomy Management System and modelled its constituents. Last but not least we have implemented and experimentally tested the SenseCare Taxonomy Management System.

Table 2. Pearson R Correlation Between Big Five Aspect Scale Traits and SenseCare Classified Emotion

Pearson R	*Openness to Exp.*	*Openness*	*Intellect*	*Conscientiousness*	*Orderliness*	*Industriousness*	*Extraversion*	*Assertiveness*	*Enthusiasm*	*Agreeableness*	*Politeness*	*Compassion*	*Neuroticism*	*Volatility*	*Withdrawal*
ML Anger	0.15	0.09	0.17	0.15	0.09	0.17	0.15	0.09	0.17	0.20	0.18	0.13	0.15	0.09	0.17
ML Fear/Anxiety	0.07	0.04	0.08	0.07	0.04	0.08	0.07	0.04	0.08	-0.15	-0.04	-0.18	0.07	0.04	0.08
ML Sadness	-0.14	-0.06	-0.18	-0.14	-0.06	-0.18	-0.14	-0.06	-0.18	-0.34	-0.19	-0.32	-0.14	-0.06	-0.18
ML Surprise	-0.12	-0.13	-0.07	-0.12	-0.13	-0.07	-0.12	-0.13	-0.07	0.26	0.21	0.18	-0.12	-0.13	-0.07
ML Joy	0.16	0.09	0.18	0.16	0.09	0.18	0.16	0.09	0.18	-0.08	-0.16	0.04	0.16	0.09	0.18
ML Disgust	-0.18	-0.10	-0.20	-0.18	-0.10	-0.20	-0.18	-0.10	-0.20	0.05	0.01	0.06	-0.18	-0.10	-0.2
ML Neutral	-0.16	-0.19	-0.06	-0.16	-0.19	-0.06	-0.16	-0.19	-0.06	0.37	0.25	0.31	-0.16	-0.19	-0.06

Table 3. Pearson R Correlation Between Self-Reported Emotions and SenseCare ML detected emotions

SR Emotion	**ML Anger**	**ML Fear / Anxiety**	**ML Sadness**	**ML Surprise**	**ML Joy**	**ML Disgust**
Fear	-0.34	0.08	0.04	0.19	0.14	-0.09
Anger	-0.03	-0.25	0.13	0.07	0.11	-0.06
Joy	0.24	-0.01	0.17	-0.09	0.07	0.06
Anxiety	-0.04	0.16	-0.12	0.13	0.17	-0.20
Sadness	0.11	-0.30	0.10	0.01	-0.04	0.09
Surprise	-0.12	-0.04	-0.20	0.36	-0.04	-0.10
Disgust	0.05	-0.01	-0.12	-0.07	0.12	-0.07

E. Requirements Analysis

Participants were requested to fill out a personality questionnaire and several questionnaires that assessed the basic emotions of Fear, Anger, Joy, Anxiety, Surprise, Disgust and Sadness (see Table 3). This meant that a raw score for each of the Big Five Personality Traits and their associated sub-traits was recorded along with a self-reported emotion after each emotional stimulus. The Personality Questionnaire consisted of 100 statements, with the participant judging how consistent the statement was to their behaviour and thoughts, that were answered on a Likert-scale from 1 to 7, with 1 representing “not like me at all” to 7 representing “very much like me”. The self-reported emotional questionnaire was also a Likert-scale ranging from 1-5. Participants were asked to report to what extent did they experience each of the basic emotions during the last video. A score of 1 represented “not at all” and 5 represented “an extreme amount”. This data was exported via excel and processed using *R*. After the study, each of the recorded videos were analysed with the SenseCare ML emotion detector for emotional analysis and classification. The sensitivity of the emotion detector was adjusted to its least sensitivity score in order to avoid Type 2 errors. A frame-by-frame emotional analysis was carried out by the software from each video footage captured from the study. The analysis per video was automatically saved to a text file with a timestamp for each recorded emotion along with the emotion classification. For later reuse of created information, all these reports must be managed by the SenseCare system in order to keep analysis results reusable for later validation. A comprehensive data management support, allowing later comprehensive description and access to this data, is needed. We concluded that first, a Taxonomy Editor is required in order to support crowdsourcing in taxonomy development and management. Second, a Category Manager is required to describe structure of a taxonomy including terms and their relationships. Thirdly, a Version Control is required to support taxonomy evolution in the taxonomy management process. Forth, a Content Rating is required to support crowd voting process of a taxonomy. Last but not least, a Categorization functionality is required in order to support content classification, faceted browsing, and system navigation.

F. Conceptual Modelling

The Taxonomy Editor lets users manage their taxonomies. Users can interact with the user interface to create, edit, delete, import, and export their taxonomies. Administrators can modify all taxonomies in the system. Only they can set a taxonomy to become the system’s seed (main) taxonomy, which can be used for classification and searching of content. Visitors to the portal can see the metadata of a taxonomy and have the ability to export the taxonomy to different formats. The module consists of one controller *Taxonomy Editor* and one model *Taxonomy*. The model describes the taxonomy by its metadata, such as e.g. title, description, authors. The controller *Taxonomy Editor* creates a new taxonomy and calls *Content Service* to create a new record in table *Registration*. Model *Registration* in the component *Content Manager* works as a public directory for all content in the KM-EP. Each content has a unique external ID stored in *Registration*. Module *Content*

Display translates this ID into internal ID and type of content, so it can be found quickly. Furthermore, *Registration* also saves common content metadata, such as e.g. image, owner, created date. Data from *Taxonomy* and *Registration* is indexed by *Solr Service* to a *Solr server*, which is a high-performance search server. User search requests are sent to this server to be processed, and the results are returned to the user in just milliseconds.

The Category Manager module enables users to manage the structure of a taxonomy. Users can create, edit, delete terms in a taxonomy and their relationships. A taxonomy has at least one term. The model *Category* describes a term of a taxonomy. A term, depending on the current ontology, can have properties, e.g., *skos:prefLabel*, which describes the preferred label of this term in SKOS. The model *Property* stores all properties of each term. Field *name* saves the name of the property and field *value* saves its value. A term also has relationships with other terms. The model *Relation* describes the relationship between terms in this taxonomy. Field *name* stores the name of the relationship, e.g., *skos:broader* which describes the term *A* as a broad term of term *B* [7].

A version control system needs to be implemented to support the taxonomy evolution process. This module lets users take a snapshot of a taxonomy, merge a current taxonomy into its origin, view the snapshot history, and rollback to a chosen version. Visitors can also view the commit history of a taxonomy whilst they browse its content display page. Here, we chose to implement a solution based on the principle of the distributed version control system Git [25]. Nevertheless, we do not deal with directories and files like in Git, but work with taxonomies, their categories, and category properties and relations. Hence, changing is needed to reflect these differences. The commit object is stored in the table *Version* with its metadata. Tree and blob objects are stored in the same table *Blob*, because a taxonomy contains only one kind of content, which is category. Git is different, where a repository stores both directories and files. Each record in table *Blob* contains both metadata of the category, its content, and checksum (hash).

In addition, taxonomy generation and evolution should be supported by giving users the ability to vote for each content including taxonomies. Hence, module Content Rating was implemented. When a user rates content, an entry is inserted into table *Rating* with the score, the ID of the user who rated the content, and the ID of the record in table *Registration* that describes the content. This record points to the taxonomy in table *Taxonomy* in case a user votes for a taxonomy. The average score of the content should be re-calculated each time a new user rates it or a user changes his/her rating for the content.

Categorization lets users categorize content into different taxonomy categories. The module introduces a new table *Categorization*. This table stores the ID of the content that was categorized, along with the ID of taxonomy, which the content was categorized into. The categorization is indexed by the search server and can be retrieved later using the searching page. The indexed categorizations are used for supporting faceted search in the future, which is an important feature and is used by many key websites, such as e.g. Amazon, eBay, Google to retrieve information. Faceted search helps to address the weaknesses of conventional search approaches and provides more effective information-seeking support to users [26].

G. *System Implementation*

The prototype, called "Taxonomy Manager", was developed as a component of the KM-EP. The implementation uses the Symfony Framework [27], which is currently one of the leading PHP Frameworks [28] supporting the creation of web applications. It has a large community, many reusable components, and high-quality documentation. The relational database system – MySQL is used for storing taxonomy metadata, properties, and relationships as well as version metadata and snapshots. At the front-end, the taxonomy tree is built with the help of JavaScript libraries – jsTree and jQuery JavaScript's framework AngularJS create a seamlessly experience for the users. Data is sent to the back-end using AJAX [29] calls. Hence, users can interactively modify data without reloading the page.

The first module of the prototype is the Taxonomy Editor. This module lets users manage their taxonomies. Users interact with the User Interface to create, edit, delete, import, and export their own taxonomies. The Taxonomy Editor also provides users with the ability to import a taxonomy from a file and insert it into the chosen taxonomy. The front-end sends the upload file's content to the back-end and receives an answer notifying if the action was successful or not. Users can also export a taxonomy or a part of it to a file and download it to their computer. After a user completed work on a taxonomy, details about it, such as e.g. metadata, structure, and version history are published. Visitors can access all of this information, export the taxonomy to an ontology file, and rate the taxonomy themselves. They also have to possibility to share the taxonomy on the social media.

The second module of the prototype is the Category Manager. This module lets users manage the structure of taxonomies. Users create, edit, delete terms of taxonomies and their relationships. The Category Manager module splits into back-end and front-end, where the front-end is implemented based on the AngularJS Framework due to its ability to update the view automatically whenever the data, which is binding to it, changes. The JavaScript library jsTree displays the taxonomy in tree format. The back-end is implemented based on the Symfony Framework. The flow of the Category Manager is as follow: When a user edits a taxonomy, the *DisplayController* in AngularJS calls API *load* in the backend. The module Category Manager reads the taxonomy from 2 tables *category* and *category_relation* in the database for the taxonomy terms and their relationships. Because the taxonomy is by default represented in a tree format, a taxonomy tree is built and cached using the distributed memory object caching system MemCached [30]. The purpose of caching the built tree is to reduce the time and resource needed from the server. Next time when a user accesses this taxonomy again, the data can be read directly from the cache.

The next module of the prototype is Version Control. This module lets users save (commit) the current state of a taxonomy, check the history of previous commits, reset (rollback) a taxonomy to a previously committed state or replace (merge) a taxonomy with one of its clones. Whilst building the taxonomy tree, a commit cache is built. The data related to each node in the taxonomy is stored. These can be the node's unique identifier, metadata, data, such as properties and relations, and hash calculated from that data. We call them a blob. A snapshot is made of many blobs. The hash will be compared with the hash of the last commit of the same node to check if there is a change in this node since the last commit. If there is no change with the node, it is not necessary to store the data again since we can always get the same data from the last commit. A user rarely makes too many changes in one version. Hence, this commit mechanism uses less space compared to saving the whole taxonomy. By checking if data was stored, we can also build the list of changes for each version as part of a commit's history.

The Content Rating module was implemented to serve the evolution process of taxonomies. The module lets users rate a content by selecting the number of stars in the user interface. It also shows the current rating of content by coloring the stars and displaying the number of users who rated the content. Figure 2 shows the activities of the Content Rating module.

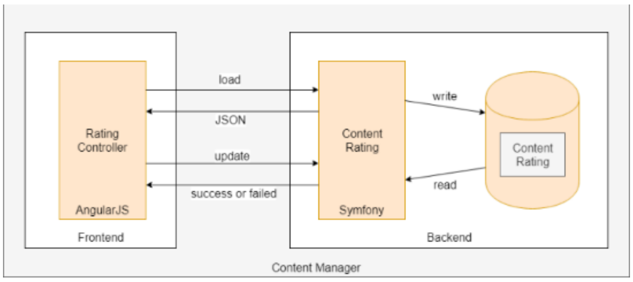

Fig. 2. Activities of the Content Rating module

Each time a visitor accesses the content display page, the AngularJS controller *RatingController* sends a *load* request to the back-end. The Symfony controller *ContentRating* in the back-end will search for all ratings of the content in table *content_rating* and calculate the average score. The calculated score is returned to the front-end in JSON format. The AngularJS controller then highlights the stars based on the score as well as showing the number of ratings for the content. Users can rate content by selecting how many stars they want. The front-end will send an *update* request with the selected score, user ID, and content ID to the back-end. The Symfony controller *ContentRating* receives the request and attempts to insert the record in the database. Furthermore, the controller also recalculates the average score of the content and sends the new score to the search server. It facilitates showing the up-to-date average score of content in the search result. The search server in this case does not need to calculate the average score of each content that belongs to the result, which saves time and resources.

Categorization is the fourth module of the prototype. This module lets users categorize content to categories, which are terms of one or different taxonomies. As mentioned above, an administrator can enable a taxonomy to be used for content classification and searching. The chosen taxonomies are called "seed taxonomy" or "base taxonomy". The keyword "base taxonomy" is automatically added into the list of keywords of the taxonomies, so the user knows which taxonomies are being used for content classification and searching. When a user clicks on the *Assign* button, the front-end sends a request *update* to the back-end with the payload, which are the list of selected categories, the ID of selected taxonomy and the ID of the editing content. The controller in the back-end reads the parameters in the payload, tries to insert the new data into the database, and notifies the front-end if it was successful or not. The backend controller also notifies the search server for this change in the categorization of the content. Hence, indexed data can be updated immediately. When searching for content, users can select category terms to filter only the content that were categorized with these terms. The numbers in the right side of terms indicate the number of content in the result list that were categorized with those terms. With the faceted feature, the user can have an overview of the classification of content in real-time and quickly find the result by selecting only relevant categories. Furthermore, this enables navigation using the taxonomy's hierarchical structure.

IV. DISCUSSION

In terms of supporting the complex and interdisciplinary knowledge domain of AC in SenseCare, the taxonomy management system achieved three goals. (1) The first goal is the development and management of initial emotion taxonomies. Hence, several taxonomies were imported to the Taxonomy Manager. Examples are the Sentient 26 Emotional Taxonomy, which is an emotional motivation framework for understanding consumer behaviour, and parts of the WHO's ICD-10 classification, which classify mental disorders. Common taxonomies like these two exemplar ones would allow sharing and comparing information easier by offering standard vocabularies and formats. (2) Furthermore, these two taxonomies along with others from different knowledge domains are used to classify scientific content of the SenseCare AC domain stored in the KM-EP's digital library, such as e.g. publications, multimedia, and person with dementia records. As a result, the content can easily be managed and found, which is the second goal of the Taxonomy Manager in the SenseCare KM-EP. (3) Finally, the analysis results produced by the emotion detection platform [1] will also have to be indexed using similar taxonomies from the Taxonomy Manager. To achieve this, analysis results like those displayed in Table 1 have to be imported as a type of scientific content and then classified in the same way as the relevant scientific publications of SenseCare's AC domain. This demonstrates that the Taxonomy Manager not only can be used to collect, classify, and provide access to materials of the initial emotion analysis and its results but also supports the work of psychology experts in a follow up study aiming at training machine learning components to classify personality traits from vectors of initial emotion classification features. This work could be much more costly without the classification, annotation, and access support of the Taxonomy Manager in the SenseCare KM-EP supporting scientific research in the domain of AC.

V. CONCLUSION

In contrast with AC, which is a relatively recent area of computer science, taxonomy is an old science that has existed for thousands of years. In modern technology, taxonomy introduced as a way to combat information overload by mean of classifying knowledge. However, developing and maintaining taxonomy is not a trivial task. Previously, it was usually completed by knowledge workers and experts, which is prone to error and costly. The objective of the work reported in this paper is to develop a taxonomy management system to support the development, evolution, and management of taxonomies. These taxonomies in turn serve machine learning emotion platforms like SenseCare in the task of emotion classification. Here, we have described the concept of taxonomy management and discussed some of the popular systems. We also introduced a new taxonomy management system. Based on this conceptual model, a prototype was implemented in the context of project SenseCare. We have also described the results of a study that examined the relationship between personality and emotions using SenseCare software. It is clear that personality and emotions are linked. Whilst the exact nature of this relationship is still yet unclear, this study has provided some insights about potential relationships. Not only that, but the divergences between the relationship between personality and self-reported emotional expressions and automated emotional expressions provokes some interesting questions. Namely, whether the use of automated emotional classification may open up the door to understanding aspects of personality and emotion that people are unable to accurately report themselves, possibly due to the pressure of answering questionnaires in a manner that reflects well on the person. Overall, the statistical analysis found several large, by research standards, effect sizes between personality traits and emotions. This is despite the small sample size of the study. The Taxonomy Manager developed in the context of SenseCare gives AC research the support it needs to monitor, diagnose, and care for people with dementia.

ACKNOWLEDGMENT

This publication has been produced in the context of the SenseCare project. This project has received funding from the European Union's H2020 Programme under grant agreement No. 690862. However, this paper reflects only the authors' view and the European Commission is not responsible for any use that may be made of the information it contains.

REFERENCES

[1] M. Healy, R. Donovan, P. Walsh and H. Zheng, "A Machine Learning Emotion Detection Platform to Support Affective Well Being," in *IEEE International Conference on Bioinformatics and Biomedicine*, 2018.

[2] "Sensor Enabled Affective Computing for Enhancing Medical Care," 19 April 2017. [Online]. Available: https://cordis.europa.eu/project/rcn/199563/factsheet/en. [Accessed 27 August 2019].

[3] F. Engel, R. Bond, A. Keary, M. Mulvenna, P. Walsh, H. Zheng, H. Wang, U. Kowohl and M. Hemmje, "SenseCare: Towards an Experimental Platform for Home-Based, Visualisation of Emotional States of People with Dementia," in *Advanced Visual Interfaces. Supporting Big Data Applications*, 2016.

[4] M. Healy and P. Walsh, "Detecting demeanor for healthcare with machine learning," in *IEEE International Conference on Bioinformatics and Biomedicine*, 2017.

[5] R. Donovan, M. Healy, H. Zheng, F. Engel, B. Vu, M. Fuchs, P. Walsh, M. Hemmje and P. M. Kevitt, "SenseCare: Using Automatic Emotional Analysis to Provide Effective Tools for Supporting Wellbeing," in *IEEE International Conference on Bioinformatics and Biomedicine*, 2018.

[6] O. P. John and S. Srivastava, "The Big Five Trait Taxonomy: History, Measurement, and Theoretical Perspectives," in *Handbook of personality: Theory and research*, New York, Guilford, 1999, pp. 102-138.

[7] B. Vu, J. Mertens, K. Gaisbachgrabner, M. Fuchs and M. Hemmje, "Supporting Taxonomy Management and Evolution in a Web-based Knowledge Management System," in *HCI 2018*, Belfast, UK, 2018.

[8] CORDIS, "Education for Data Intensive Science to Open New science frontiers," 24 June 2019. [Online]. Available: https://cordis.europa.eu/project/rcn/198292/factsheet/en. [Accessed 11 September 2019].

[9] B. Vu and M. Hemmje, "Supporting Taxonomy Development and Evolution by Means of Crowdsourcing," in *International Conference on Knowledge Engineering and Ontology Development*, Wien, 2019.

[10] C. G. DeYoung, "Cybernetic big five theory," *Journal of Research in Personality*, vol. 56, pp. 33-58, 2015.

[11] L. C. Quilty, F. D. Fruyt, J.-P. Rolland, S. H. Kennedy, P. Rouillon and R. M. Bagby, "Dimensional personality traits and treatment outcome in patients with major depressive disorder," *Journal of Affective Disorders*, pp. 241-250, 2008.

[12] G. Blickle, "Personality traits, learning stratigies, and performance," *European Journal of personality*, pp. 337-352, 1996.

[13] D. Bruno and H. Richmond, "The truth about taxonomies," *Information Management Journal*, p. 44, 2003.

[14] I. Hunink, R. van Erk, S. Jansen and S. Brinkkemper, "Industry Taxonomy Engineering: the Case of the European Software Ecosystem," in *Proceedings of the Fourth European Conference on Software Architecture: Companion Volume*, 2010.

[15] M. C. Daconta, L. J. Obrst and K. T. Smith, The Semantic Web: A guide to the future of XML, Web Services, and Knowledge Management, John Wiley & Sons, 2003.

[16] O. P. John, L. P. Naumann and C. J. Soto, "Paradigm shift to the integrative big five trait taxonomy," *Handbook of personality: Theory and research*, pp. 114-158, 2008.

[17] B. R. Little, Me, myself and us: The science of personality and the art of well-being, New York: Public Affairs, 2014.

[18] D. P. McAdams, The art and science of personality development, Guilford Publications, 2015.

[19] C. G. DeYoung, L. C. Quilty and J. B. Peterson, "Between facets and domains: 10 aspects of the Big Five," *Journal of personality and social psychology*, p. 880, 2007.

[20] J. B. Hirsh, C. G. DeYoung, X. Xu and J. B. Peterson, "Compassionate liberals and polite conservatives: Associations of agreeableness with political ideology and moral values," *Personality and Social Psychology Bulletin*, pp. 655-664, 2010.

[21] B. Walli, "Taxonomy 101: The Basics and Getting Started with Taxonomies," 15 August 2014. [Online]. Available: http://www.kmworld.com/Articles/Editorial/What-Is/Taxonomy-101-The-Basics-and-Getting-Started-with-Taxonomies-98787.aspx.

[22] "Taxonomy Thesaurus Manage," [Online]. Available: https://www.poolparty.biz/taxonomy-thesaurus-management.

[23] Biomedical Informatics Research Group, "WebProtégé," [Online]. Available: https://webprotege.stanford.edu/. [Accessed 16 September 2019].

[24] T. Tudorache, C. Nyulas, N. F. Noy and M. A. Musen, "WebProtégé: A Collaborative Ontology Editor and Knowledge Acquisition Tool for the Web," *Semant Web*, pp. 89-99, 2013.

[25] "Git," n.d.. [Online]. Available: https://git-scm.com/.

[26] D. Tunkelang, Faceted Search, Morgan & Claypool, 2009.

[27] "Symfony Framework," n.d.. [Online]. Available: http://symfony.com/.

[28] "What is PHP?," [Online]. Available: https://php.net/manual/en/intro-whatis.php.

[29] "Ajax," n.d.. [Online]. Available: https://developer.mozilla.org/en-US/docs/Web/Guide/AJAX.

[30] "Memcached," n.d.. [Online]. Available: https://memcached.org/.