Comparing automatically detected reflective texts with human judgements

Conference or Workshop Item

Comparing Automatically Detected Reflective Texts with Human Judgements

Thomas Daniel Ullmann⋆, Fridolin Wild, and Peter Scott

Knowledge Media Institute, The Open University
Walton Hall, MK7 6AA Milton Keynes, United Kingdom
{t.ullmann,f.wild,peter.scott}@open.ac.uk
http://kmi.open.ac.uk

Abstract. This paper reports on the descriptive results of an experiment comparing automatically detected reflective and not-reflective texts against human judgements. Based on the theory of reflective writing assessment and their operationalisation five elements of reflection were defined. For each element of reflection a set of indicators was developed, which automatically annotate texts regarding reflection based on the parameterisation with authoritative texts. Using a large blog corpus 149 texts were retrieved, which were either annotated as reflective or not-reflective. An online survey was then used to gather human judgements for these texts. These two data sets were used to compare the quality of the reflection detection algorithm with human judgments. The analysis indicates the expected difference between reflective and not-reflective texts.

Keywords: reflection, detection, thinking skills analytics

1 Introduction

The topic of reflection has a long-standing tradition in the area of educational science as well as in technology-enhanced learning. Reflection is seen as a key competency. These are competencies, which are important for society, to help meeting important demands for all individuals and not only for specialists. Reflection is at the "heart of key competencies" for a successful life and a well-functioning society [25].

The focus of this research is on reflective writings. A reflective writing is one of many ways to manifest the cognitive act of reflection. Common forms are diaries, journals, or blogs, which serve a person as a vehicle to capture reflections.

Although reflection has been present in the modern educational discourse since at least 1910 [11], methods for the assessment of reflective writings are a relatively recent development. They are not in their infancy, but they are neither fully established. Wong et al. [37] states that there is a lack of empirical research on methods of how to assess reflection, and that the discussion is more driven by

⋆ Corresponding author

theorising concepts of reflection and its use. Plack et al. [27, p. 199] more recently states "(...) yet little is written about how to assess reflection in journals".

Classical tools to identify evidence of reflections are questionnaires (e.g. [1, 3]), and manual content analysis of reflective writings (for an overview see Dyment and O'Connell [12]). These methods are time-consuming and expensive. Due to their nature, the evaluations of reflective writings and feedback are usually available far after the act of writing, as it first has to be processed by an expert. In addition, due to the personal nature of reflection some people prefer not to share them, although feedback would benefit their reflective writing skills.

The automated detection of reflection is a step forward to mitigate these problems, as well as it provides a new perspective on the research of reflection evaluation methods.

As a first step towards this goal, text was annotated and based on the annotation rules were defined. These rules mapped five elements of reflection. Then the reflection detector was parameterised based on authoritative texts. This baseline parameterisation was used to distinguish texts that fulfilled the rule criteria and afterwards referred to as reflective texts, and texts, which do not satisfy these criteria, referred to as not-reflective. A larger blog corpus was automatically analysed. The annotated texts were rated by human judges. This paper reports the results of the comparison between automated detection of reflection and human ratings.

2 Situating the Research in the Research Landscape

The automated detection of reflection is part of the broader field of learning analytics, especially social learning content analysis [13].

Two related prominent approaches for identifying automatically cognitive processes have emerged in the past. The first approach draws from the associative connection between cue words and acts of cognition. This approach explicitly uses feature words associated with psychological states. Pennebaker and Francis [26], for example, developed the Linguistic Inquiry and Word Counting tool to research the link between key words and its impact on physical health and academic performance using a bank of over 60 controlled vocabularies in the detection of emotion and cognitive processes. Bruno et al. [6] describe an approach for analysing journals using a mental vocabulary. This semi-automatic approach focuses on the detection of cognitive, emotive, and volitive words, enabling them to highlight changes in the use of these mental words over a course term. Chang and Chou [7] are using a phrase detection system to study reflection in learners' portfolios. The system serves as a pre-processor of contents, thereby emphasising specific parts-of-speech (in their case: stative verbs in Mandarin), which then later helped experts to assign the automatically annotated words to four categories associated with reflection, labelled as emotion, memory, cognition, and evaluation.

The second type of approaches relies on probabilistic models and machine learning algorithms. McKlin [21] describes an approach using artificial neural networks to categorise discussion posts regarding levels of cognitive presence. The concept of cognitive presence reflects according to Garrison et al. [14, p. 11] "(...) higher-order knowledge acquisition and application and is most associated with the literature and research related to critical thinking". Cognitive presence consists of four categories: triggering events, exploration, integration, and resolution. The cognitive presence model was also used in the ACAT system [8]. In this system, a Bayesian classifier was used to distinguish content according to the four categories of the cognitive presence model. Rosé et al. [28] describe the use of a set of classification algorithms (Naïve Bayes, Support Vector Machines, Decision Trees) to automatically annotate sentences from discussion forums related to - amongst others - epistemic activity, argumentation, or social regulation.

3 Research Question

The wider goal of this research is to evaluate the boundaries of automated detection of reflection. This includes the question of to what extent it is possible to algorithmically codify reflection detection that validly and reliably detects and measures elements and depth of reflection in texts and how these results compare to human judgements. This is an on-going research process. Within this paper the focus lies on the following questions:

1. How does automated detection of reflection relate with human judgments of reflection?
2. What are reasonable weights to parameterise the reflection detector?

Regarding the first question the goal is to compare automatically detected reflective texts with texts that do not satisfy the criteria of a reflective text, with human judgments. It is expected that the two categories will differ. The second question refers to the weights of the reflection detection of each element of reflection. Based on a set of reflective texts weights will be determined. It is expected that by using these weights, the reflection detector will find reflective texts, which are also marked as reflective by human judges.

4 Elements of Reflection

Up to now, an agreed model of reflection does not exist. This might be due to the variety of contexts, in which reflection research is embedded (e.g. medical area, psychology, vocational education). With this, certain elements of reflection are more important in a given context than in others contributing to this variety.

It seems however, that there are certain repeating elements of reflection, which will build the foundation of the model used in this paper. The elements presented here are based on the major streams of the theoretical discussion on reflection.

The elements of reflection used in this paper are the following:

1. Description of an experience: This element of reflection sets the stage for it. It is a description of what was happening. Boud et al. [4, p. 26] describes it as returning to experience by recapturing the most important parts of the event. The writer is recalling and detailing the salient moments of the event. The description of the happening can be either the description of external events as the source of reflection, but also descriptions of the inner situation of the person, for example their thoughts or emotions. There can be many themes, which were the reason or trigger of the writer to engage in reflective writing. Some common themes are the following.
   - Conflict: A description of an experienced conflict (either a conflict of the person with him/herself or with another person/s or situations). The conflict can be presented as a disorienting dilemma, which is either solvable or on-going.
   - Self-awareness: Recognising that cognitive or emotional factors as a driving force of own beliefs and that these beliefs are shaping own actions.
   - Emotions: Feelings are frequently cited as a starting point of reflection. As with the other topics emotions might be part of a reflection but they are not necessarily part of every one of them [24, p. 88]. Boud et al. [4, p. 26] emphasises to use helpful feelings and to remove or to contain obstructive ones, as a goal of a reflection. It can be seen as a reaction to a personal concern about an event. Dewey [10, p. 9] states that the starting point of a reflection can be a perceived as the perplexity of difficulty, hesitation or doubt, but also something surprising, new, or never experienced before.
2. Personal experience: As reflection is about own experiences, one might expect that they are self-related, and ought to tell a personal experience. Although it seems convincing that reflective writing should be about own experiences, there still exists a certain debate. Moon [24, p. 89] argues reflective writing does not necessarily needs to be written in first person. However, in the case of a deep reflection, the writer often expresses self-awareness of individual behaviour using the first person perspective. Hatton and Smith [15] describe it as an inner dialogue or monologue that forms part of the dialogic reflection of their reflection model. Boyd and Fales [5] call it personal or internal examination and Wald et al. [36] emphasis on the existence of the own voice expressed in the writing, indicating that the person is fully present.
3. Critical analysis: Mezirow [22] states that the critical questioning of content, of process, and premises of experiences in order to correct assumptions or beliefs, might lead to new interpretations and new behaviour. Dewey [10, pp. 118, 199-209] speaks of the importance of testing of hypothesises by overt or imaginative action. It is this critical analysis, which helps the writer to step back from the experience in order to be able to mentally elaborate or critique own assumptions, values, beliefs, and biases. This process of mulling over or mental elaboration can contain an analysis, synthesis, evaluation of experience, testing or validation of ideas, argumentation and reasoning, hypothesising, recognising inconsistencies, finding reasons or justifications for own behaviour or of others, linking of (association) and integrating ideas.

4. Taking perspectives into account: The frame of reference can be formed in the dialogue with others, by comparing reactions with other experiences, but also by referring to general principles, a theory, or a moral or philosophical position [33]. A change of perspective can shed new insights, and helps to reinterpret experience [22].
5. Outcome of the reflective writing: According to Wald et al. [36] a reflection can have two outcomes: Either the writer arrives to new understanding (transformative learning) or at confirmatory learning (meaning structures are confirmed). Both touch the dimension of reflection-for-action [17]. The outcome of a reflection is especially important in an educational context. It sums up what was learned, concludes, sketches future plans, but might also comprise a sense of breakthrough, a new insight and understanding.

While these elements are presented separately, there is still an overlap between them. For example, the description of an experience can already be critical and contain multiple perspectives. Wong et al. [37] subsume validation, appropriation and outcome of reflection as part of perspective change, while Wald et al. [36] puts meaning making and critical analysis into one category.

These five elements of reflection build the foundation of the theoretical framework. For each element a set of indicators was developed. Each indicator is mapped back into the elements of reflection using a set of rules. These rules define the relation or mapping between the indicators and the element of reflection.

5 Reflection Detection Architecture

With the help of several analysis engines that wrap linguistic processing pipelines for each classifier, elements of reflection can be annotated. The analysis component is then used to aggregate overviews informing about the level of reflection identified. For an overview of the architecture, see Ullmann [35].

5.1 Description of the Annotators

A set of annotators has been developed. Each annotation consists of its own type and can have one or more features. An annotation can span over a text from single characters, to words, to sentences, or even the whole text. For this paper, the following annotators were used.

– NLP annotator: The NLP annotator makes use of the Stanford NLP parser [9, 18, 34]. It is used to annotate part-of-speech, sentences, lemma, linguistic dependency, and co-references.
– The premise and the conclusion annotator use a handpicked selection of keywords indicating a premise (e.g. assuming that, because, deduced from) or conclusion (e.g. as a result, therefore, thus).
– The self-reference annotator is based on keywords referring to the first person singular (I, me, mine, etc.), while the "pronoun other" annotator contains keywords referring to the other/s (he, they, others, someone, etc.).

- The reflective verb annotator is a refined version of Ullmann [35], making use of reflective verbs (e.g. rethink, reason, mull over).
- The learning outcome annotator is based on Moon [23, pp. 68-69] (lemmas: define, name, outline, etc.), while the Bloom [2] taxonomy annotator contains keywords for the categories "remember", "understand", "apply", "analyse", "evaluate", and "create".
- The future tense annotator is built from a selected list of key words, indicating future tense (will, won't, ought, etc.).
- The achievement, causation, certainty, discrepancy, and insight annotator are based on the LIWC tool [26], but refined and based on lemmas.
- The surprise annotator contains a refined set of nouns, verbs, and adjectives from the SemEvalTask[1] [31], which in turn are based on WordNet affect [32].

5.2 Description of the Analysis Component

While the analysis of the annotators can already help to gain insights regarding the reflectivity of the text, the aggregation of annotators adds an additional layer of meaning. Besides UIMA as a framework to orchestrate the annotators, the Drools framework - especially its rule engine - was leveraged to infer knowledge from the annotations. This has several benefits starting from the ability to infer new facts, chain facts from low-level facts to high-level constructs, to update facts, and to reject facts. The rules are expressed in IF - THEN statements (for example, if A is true then B).

As a simplified example (see 5.2) I show three rules to infer whether a sentence shows evidence of personal use of the reflective verb vocabulary (the rule is described in natural language and not using the notation of Drools). This is one of the six rules of the indicator critical analysis.

Listing 1.1. Rule example

```
FOR ALL sentences of the document:
IF sentence contains a nominal subject
AND IF it is a self-referential pronoun
AND IF the governor of this sentence is contained in the
    vocabulary reflective verbs
THEN add fact "Sentence is of type personal use of reflective
     vocabulary"
```

For each element of the reflection a set of rules can be used to describe the mapping between the annotations and the element of reflection. The high-level rules of each element are then combined to a rule/s, which indicates reflection or grades of reflection. The micro level of analysis is the set of facts formed by the annotations, the meso level represents the set of rules for each element, and the macro level is the set of rules indicating the high-level construct (in this case reflection).

[1] http://www.cse.unt.edu/ rada/affectivetext/

6 Method

The discussion of the method will follow two strands. First, we will outline the method used to distinguish texts regarding their reflective quality using the reflection detector. This includes the mapping of indicators to the elements of reflection and the parameterisation of the macro rule to detect reflection. The result of the automatic classification labels each text with either "reflective" or "not-reflective". The second strand describes the method used to gather the human judgments using on an online questionnaire.

6.1 Assignment of Indicators to Elements of Reflection

This experiment uses 16 rules, which indicate a facet of an element of reflection. For each element of reflection, a set of indicators was designed. The development of each indicator was an iterative process. Based on the experience of the first author with reflective texts several versions of each indicator were developed, and the most promising ones were kept. Each indicator was tested with sample texts, including reflective texts, not-reflective ones, and self-generated test cases. The goal of this approach was to generate sound indicators, which could then be tested against empirical data.

Altogether 28 rules form the meso-level. Several of these rules are chained together, leaving 16 rules at the end of the chain. These 16 rules were assigned to each of the five elements of reflection based on the elements derived from theory (see Table 1).

Elements of reflection	Indicators (based on rule inference)
Description of an experience	Past tense sentence with self-related pronoun as subject. Present tense sentence with self-related pronoun as subject. Sentence with surprise keyword and self-related pronoun as subject.
Personal experience	All indicators, which are based on self-related pronouns. Question sentence, in which the subject is a self-related pronoun.
Critical analysis	Sentences with premise, conclusion, and causation keywords. Sentences with certainty or discrepancy as keyword and using as subject a self-related pronoun. Sentences, which have a self-related pronoun as subject and a reflective verb as governor.
Sentences that take other perspectives into account	Sentences, which have a "pronoun others" as subject and a self-related pronoun as object. Sentences, which have a self-related pronoun as subject and pronouns others as object.

Outcome	Sentences, which have self-related pronoun as subject and keywords coming from the Bloom [2], or Moon [23, pp. 68-69] taxonomy of learning outcomes. Sentences, which have a self-related pronoun as subject and a keyword expressing insights. Future tense sentences with self-related pronoun as subject.

Table 1: Mapping of elements of reflection to indicators (as a self-related pronoun we understand a 1st person singular pronoun, while a pronoun referring to others is termed as pronoun other).

According to this mapping, sentences, which are personal and written in the past or present, or contain surprise, belong to the element "description of experience". The element of "personal experience" is implicitly covered by all sentences, which are self-related. Additional self-related questions are covered. Sentences with premise, conclusion, causation, certainty, discrepancy, or reflective key words are subsumed in the element "critical analysis". "Taking perspectives into account" uses two rules, while the "outcome" dimension is based on the Moon [23, pp. 68-69] and Bloom [2] taxonomy of learning outcomes, but also insight keywords and sentences, which refer to future events.

6.2 Parameterising the Reflection Detection Architecture

One of the imminent questions is which weight should be given to each indicator to form a reflective text. In this context "how many occurrences of each indicator satisfy as criteria indicating an evidence of an element of reflection?" To parameterise the reflection detection analytics component 10 texts found in the reflection literature marked as prototypical reflective writings were used. This reference corpus contains 10 texts taken from the instructional material of Moon [24], and the examples of the papers of Korthagen and Vasalos [19], and Wald et al. [36] supplemental material. The texts were automatically annotated and analysed. For each element of reflection the individual indicators were aggregated and the arithmetic mean calculated. The results are broken down in the following table (see Table 2).

Elements of reflection	Mean
Description of an experience	5.23
Self-related questioning (several other indicators implicitly contain the element "personal experience")	0.80
Critical analysis	3.55
Taking other perspectives into account	0.45
Outcome	4.13

Table 2: Parameters for the elements of reflection.

These figures are used in the analytics component of the reflection detection engine as parameters. According to this, a text is reflective if all of the following conditions are met:

– The indicators of the "description of experiences" fire more than four times.
– At least one self-related question.
– The indicators of the "critical analysis" element fire more than 3 times.
– At least one indicator of the "taking perspectives into account" fires.
– The indicators of the "outcome" element fire more than three times.

Texts detected with these parameters belong to the group "reflective", while texts, which do not satisfy any of the conditions (fires zero times), belong to the group "not-reflective".

6.3 The Questionnaire

The aim of the design of the online questionnaire was two-fold. On the one hand, the formulation of the questions had to be suitable for a layperson audience regarding the reflection research terminology, and on the other hand to allow that, the participant could leave the survey at any time. The questionnaire consists of the following building blocks. Each page contained five blog posts. After each blog post, seven questions were displayed, which refer to the reflective quality of the blog post. Each item had a short description to clarify the task. A six-level Likert scale was used ranging from strongly agree to strongly disagree. All seven items were required.

1. The text contains a description of what was happening. Description: Does the text re-capture an important experience of the writer? This could be a description of a situation, event, inner thoughts, emotions, conflict, surprise, beliefs, etc.
2. The text shows evidence of a personal experience. Description: The text is written with an inner voice. Contains passages, which are self-related, describing an inner examination, or even contains an inner monologue/dialogue, etc.
3. The text shows evidence of a critical analysis. Description: Does the text contain an examination of what was happening? This might be an evaluation, linking or integration of ideas, argumentation, reasoning, finding justifications or inconsistencies, etc.
4. The text shows evidence of taking other perspectives into account. Description: This includes recognising alternative explanations or viewpoints, or a comparison with other experiences, also references to general principles, theories, moral or philosophical positions.
5. The text contains an outcome. Description: The text contains a description of what was learned, what is next, conclusions, future plans, decisions to take, etc. It might even contain a sense of breakthrough, new insights or understanding.

6. The text describes what happened, what now, and what next. Description: Does the text contain evidences of all three questions: What happened? What now? What next?
7. The text is reflective: Description: A reflective text shows evidences of critical analysis of situations, experiences, beliefs in order to achieve deeper meaning and understanding.

The first five items of the questionnaire reflect the above outlined elements of reflection. The description of item seven follows the definition of reflection based on Mann et al. [20]. Item six refers to the time-dependent dimensions of reflection [17, 30]: reflection-on-action, reflection-in-action and reflection-for-action.

6.4 Text Corpus

The text corpus is based on the freely available blog authorship corpus [29]: "The Blog Authorship Corpus consists of the collected posts of 19,320 bloggers gathered from blogger.com in August 2004. The corpus incorporates a total of 681,288 posts and over 140 million words - or approximately 35 posts and 7250 words per person" [29]. The blog authorship corpus was used as a vehicle to examine texts according to their reflectivity[2]. From the whole blog authorship corpus the first 150 blog files were taken and automatically analysed. A file contains all individual blog posts of one blog. Short blog posts (less than 10 sentences) and blog posts in another language than English were removed. The rational was that a reflective writing that fulfills the above outlined elements is usually a longer text. In total 5176 blog posts were annotated. In total 4.842.295 annotations were made, which resulted into 178.504 inferences. The reflection detector classified the texts, and after the removal of texts with more than three unsuitable words (all remaining bad words were replaced by a placeholder), 149 texts were detected (95 reflective and 54 not-reflective ones).

6.5 Survey Sample

The data of the survey was collected during July 2012. The set was complete in the last week of July. The questionnaire did not collect personal data. The online survey showed the blog posts together with the questions in randomised order. Each page contained five blog posts. The aim of the survey was to receive at least three complete ratings on all questions per blog posts. A small incentive was granted to each participant of the survey. In total 464 judgements were made.

In a test trial of the first author, the average time to rate each page was about six minutes, which is in line with the average duration of the participants (371 sec.). The initial analysis however revealed that several participants only spent seconds per page. To assure that at least a minimum time was spent with the

[2] as a prepared reflective text corpus is not available, which could have been used as a gold standard

task the data were filtered and judgements, which took less than 300 seconds, were eliminated. This reduced the amount of judgements to 202 (74 for the not-reflective texts and 128 for the reflective texts).

7 Results

The initial results of the experiment are summarised in Table 3. It shows for each of the two conditions the mean, the standard deviation, and the sample size. The values of the items range from 1 (strongly agree) to 6 (strongly disagree). The hypothesis is that the reflection category should have stronger agreement (smaller number) than the not-reflective category. Comparing the face value of the mean values, this tendency can be confirmed. Especially the element "personal experience" and "reflective" show a higher difference between the means. On average, more people agreed that the texts of the automatically categorised group "reflection" contain more evidence of personal experience and reflection, than the "not-reflective" group.

	reflective			notreflective		
element	N	Mean	SD	N	Mean	SD
situation	128	2.10	1.33	74	3.62	1.73
personal	128	2.11	1.43	74	3.84	1.54
critical	128	2.92	1.40	74	4.12	1.60
perspective	128	3.25	1.46	74	4.53	1.55
outcome	128	3.34	1.62	74	4.30	1.64
whatnext	128	2.71	1.43	74	4.03	1.64
reflective	128	2.51	1.48	74	4.09	1.63

Table 3. Descriptive results.

The data of this analysis is based on the average time anticipated to fulfill the task. This has the benefit of leaving most of the judgements for the descriptive analysis. The next section examines if the differences between reflective and not-reflective texts still hold, if the requirements on the dataset are taken more strictly.

The data was gathered with Amazon's Mechanical Turk. This has the major advantage, that the experiment is not influenced by the researcher and that the coders are independent from each other. However, it comes with some costs, which make a thorough analysis of the data necessary.

An inspection of the data reveals that the time spent on each page varies. Many coders spend only a few seconds on each page, which indicates that they filled in the questionnaire more or less randomly. This led to filter judgments spent less than 120 seconds.

Besides the filtering of results based on time, it was also checked if one person filled out the two pages spending exact the same time for both. Although this could happen by chance, these persons were dismissed. This pattern can arouse, if for example, a script was written, which randomly fills in the answers, waits for a certain duration and then fills in the next page with the exact same time. This suspect of data manipulation was nourished by the observed behaviour that some of the people only needed seconds to fill out a page of the questionnaire, which could mean they are answered automatically or the person randomly selects answers, and additional reports on the quality of the judgments[3]. Based on the analysis three people were dismissed.

After the removal of these judgments, the whole dataset was re-evaluated to make sure that at least two people rated each item. The initial goal was to have at least three ratings per item. However, the deletion of the judgments reduced the set to a degree, that for the experiment two ratings per item had to suffice. To compensate the benefit of additional coders the standard deviation was taken into account. If the standard deviation was bigger than 1.5, then the whole rating was discarded. This assures that only items, which were consistently rated by at least two coders remain in the dataset.

With this removal, some of the items did not have any more ratings on all seven items. These items were removed as well. The resulting descriptive statistics can be seen in the following table (Table 4).

	reflective			notreflective		
element	N	Mean	SD	N	Mean	SD
situation	18	1.87	0.66	10	3.27	0.97
personal	18	1.65	0.79	10	3.57	1.35
critical	18	2.66	0.88	10	3.52	1.34
perspective	18	3.19	1.12	10	4.37	1.13
outcome	18	2.71	0.96	10	3.42	1.44
whatnext	18	2.27	0.79	10	3.32	1.20
reflective	18	2.11	0.10	10	3.52	1.44

Table 4. Difference between reflective and not-reflective texts

The descriptive statistics of this refined analysis is in line with the results above. If a text is reflective then the human coders agree more with the asked six questions, than with less reflective texts.

8 Discussion

The results indicate that on average the two types of text not only differ within the reflection detection system, but also in the perception of human judgements. The anticipated stronger agreement of the reflective category is reflected in the

[3] http://www.behind-the-enemy-lines.com/2010/12/mechanical-turk-now-with-4092-spam.html

mean values compared to the not-reflective category. While these initial results of the analysis are already encouraging, further confirmatory testing is necessary.

The parameterisation of the reflective texts is crucial, as these values set the base line for the reflection detection. While 10 texts already give insights on the weight of each indicator a larger corpus of reflective texts would be helpful for fine-tuning the weights. The inherent problem is that by now no larger corpus of high quality reflective texts exists, which are suitable for natural language processing. The approach described here is a first step towards a reflective text corpus. The assignment of indicators to the elements of reflection is in essence an additive model. This is seen already as a good starting point, as with this simple rule already differences are detectable. However, future research will consider more complex rules, which represent the essence of reflective texts more accurate, by taking into account a wider body of reflective texts for parameterisation.

9 Outlook

Reflection is an important part in several theories and has many facets. This faceted character of reflection makes it a fascinating area of research as each element of reflection bears its own research problem, as well as aggregating indicators to a meaningful whole is yet to research. First steps have been made and some of them were sketched in this paper. Currently, the focus of this research is the development and evaluation of the analytics component of the reflection detection architecture. As a next step the data gained from this experiment, will be further analysed with the goal to refine the parameters of the reflection detector.

One possible application scenario especially useful for an educational setting is to combine the detection with a feedback component. The described reflection detection architecture with its knowledge-based analysis component can be extended to provide an explanation component, which can be used to feedback why the system thinks it is a reflective text, together with text samples as evidences.

References

[1] Aukes, L.C., Geertsma, J., Cohen-Schotanus, J., Zwierstra, R.P., Slaets, J.P.: The development of a scale to measure personal reflection in medical practice and education. Medical Teacher 29, 177–182 (Jan 2007), `http://informahealthcare.com/doi/abs/10.1080/01421590701299272`

[2] Bloom, B.S.: Taxonomy of educational objectives. Longmans, Green (1954)

[3] Bogo, M., Regehr, C., Katz, E., Logie, C., Mylopoulos, M.: Developing a tool for assessing students' reflections on their practice. Social Work Education 30, 186–194 (Mar 2011), `http://tandfprod.literatumonline.com/doi/abs/10.1080/02615479.2011.540392`

[4] Boud, D., Keogh, R., Walker, D.: Reflection: Turning Experience into Learning. Routledge (Apr 1985)

[5] Boyd, E.M., Fales, A.W.: Reflective learning. Journal of Humanistic Psychology 23(2), 99 –117 (1983), http://jhp.sagepub.com/content/23/2/99.abstract
[6] Bruno, A., Galuppo, L., Gilardi, S.: Evaluating the reflexive practices in a learning experience. European Journal of Psychology of Education 26, 527–543 (May 2011), http://www.springerlink.com/index/10.1007/s10212-011-0061-x
[7] Chang, C., Chou, P.: Effects of reflection category and reflection quality on learning outcomes during web-based portfolio assessment process: A case study of high school students in computer application courses. TOJET 10(3) (2011)
[8] Corich, S., Kinshuk, L.M.: Measuring critical thinking within discussion forums using a computerised content analysis tool. the Proceedings of Networked Learning (2006)
[9] De Marneffe, M.C., MacCartney, B., Manning, C.D.: Generating typed dependency parses from phrase structure parses. In: Proceedings of LREC. vol. 6, p. 449–454 (2006), http://nlp.stanford.edu/manning/papers/LREC_2.pdf
[10] Dewey, J.: How we think: A restatement of the relation of reflective thinking to the educative process. DC Heath Boston (1933)
[11] Dewey, J.: How we think. Courier Dover Publications (republication in 1997 of the work orginally published in 1910 by D. C. Heath & Co.) (1910)
[12] Dyment, J.E., O'Connell, T.S.: Assessing the quality of reflection in student journals: a review of the research. Teaching in Higher Education 16, 81–97 (Feb 2011), http://www.tandfonline.com/doi/abs/10.1080/13562517.2010.507308
[13] Ferguson, R., Shum, S.B.: Social learning analytics: five approaches. In: Proceedings of the 2nd International Conference on Learning Analytics and Knowledge. p. 23–33. LAK '12, ACM, New York, NY, USA (2012), http://doi.acm.org/10.1145/2330601.2330616
[14] Garrison, D.R., Anderson, T., Archer, W.: Critical thinking, cognitive presence, and computer conferencing in distance education. American Journal of distance education 15(1), 7–24 (2001)
[15] Hatton, N., Smith, D.: Reflection in teacher education: Towards definition and implementation. Teaching and Teacher Education 11(1), 33–49 (Jan 1995), http://www.sciencedirect.com/science/article/pii/0742051X9400012U
[16] Kember, D., McKay, J., Sinclair, K., Wong, F.K.Y.: A four-category scheme for coding and assessing the level of reflection in written work. Assessment & Evaluation in Higher Education 33, 369–379 (Aug 2008), http://www.tandfonline.com/doi/full/10.1080/02602930701293355
[17] Killion, J.P., Todnem, G.R.: A process for personal theory building. Educational Leadership 48(6), 14–16 (1991), http://www.eric.ed.gov/ERICWebPortal/detail?accno=EJ422847
[18] Klein, D., Manning, C.D.: Accurate unlexicalized parsing. In: IN PROCEEDINGS OF THE 41ST ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS. p. 423–430 (2003)

[19] Korthagen, F., Vasalos, A.: Levels in reflection: core reflection as a means to enhance professional growth. Teachers and Teaching: Theory and Practice 11, 47–71 (Feb 2005), http://www.tandfonline.com/doi/abs/10.1080/1354060042000337093

[20] Mann, K., Gordon, J., MacLeod, A.: Reflection and reflective practice in health professions education: a systematic review. Advances in Health Sciences Education 14, 595–621 (Nov 2007), http://www.springerlink.com/content/a226806k3n5115n5/

[21] McKlin, T.: Analyzing cognitive presence in online courses using an artificial neural network. Middle-Secondary Education and Instructional Technology Dissertations p. 1 (2004)

[22] Mezirow, J.: On critical reflection. Adult Education Quarterly 48(3), 185–198 (May 1998)

[23] Moon, J.A.: The Module & Programme Development Handbook: A Practical Guide to Linking Levels, Learning Outcomes & Assessment. Kogan Page (Mar 2002)

[24] Moon, J.A.: A handbook of reflective and experiential learning. Routledge (Jun 2004)

[25] OECD: The Definition and Selection of Key Competencies (DeSeCo): Executive Summary. OECD (2005), http://www.oecd.org/dataoecd/47/61/35070367.pdf

[26] Pennebaker, J.W., Francis, M.E.: Cognitive, emotional, and language processes in disclosure. Cognition & Emotion 10(6), 601–626 (Nov 1996), http://www.tandfonline.com/doi/abs/10.1080/026999396380079

[27] Plack, M., Driscoll, M., Blissett, S., McKenna, R., Plack, T.: A method for assessing reflective journal writing. Journal of allied health 34(4), 199–208 (2005)

[28] Rosé, C., Wang, Y.C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., Fischer, F.: Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. International Journal of Computer-Supported Collaborative Learning 3(3), 237–271 (Jan 2008), http://www.springerlink.com/content/j55358wu71846331/

[29] Schler, J., Koppel, M., Argamon, S., Pennebaker, J.: Effects of age and gender on blogging. In: Proceedings of the AAAI Spring Symposia on Computational Approaches to Analyzing Weblogs. p. 27–29 (2006), https://www.aaai.org/Papers/Symposia/Spring/2006/SS-06-03/SS06-03-039.pdf

[30] Schön, D.: Educating the reflective practitioner. Jossey-Bass San Francisco (1987)

[31] Strapparava, C., Mihalcea, R.: Semeval-2007 task 14: Affective text. Proc. of SemEval 7 (2007), http://acl.ldc.upenn.edu/W/W07/W07-2013.pdf

[32] Strapparava, C., Valitutti, A.: WordNet-Affect: an affective extension of WordNet. In: Proceedings of LREC. vol. 4, p. 1083–1086 (2004), http://hnk.ffzg.hr/bibl/lrec2004/pdf/369.pdf

[33] Surbeck, E., Han, E.P., Moyer, J.E.: Assessing reflective responses in journals. Educational Leadership 48(6), 25–27 (1991), http://www.eric.ed.gov/ERICWebPortal/detail?accno=EJ422850

[34] Toutanova, K., Klein, D., Manning, C.D., Singer, Y.: Feature-rich part-of-speech tagging with a cyclic dependency network. In: IN PROCEEDINGS OF HLT-NAACL. p. 252–259 (2003)

[35] Ullmann, T.D.: An architecture for the automated detection of textual indicators of reflection. In: Reinhardt, W., Ullmann, T.D., Scott, P., Pammer, V., Conlan, O., Berlanga, A. (eds.) Proceedings of the 1st European Workshop on Awareness and Reflection in Learning Networks. pp. 138–151. Palermo, Italy (2011), http://ceur-ws.org/Vol-790/

[36] Wald, H.S., Borkan, J.M., Taylor, J.S., Anthony, D., Reis, S.P.: Fostering and evaluating reflective capacity in medical education: Developing the REFLECT rubric for assessing reflective writing. Academic Medicine 87(1), 41–50 (Jan 2012), http://journals.lww.com/academicmedicine/Abstract/2012/01000/Fostering_and_Evaluating_Reflective_Capacity_in.15.aspx

[37] Wong, F.K., Kember, D., Chung, L.Y.F., Yan, L.: Assessing the level of student reflection from reflective journals. Journal of Advanced Nursing 22(1), 48–57 (Jul 1995), http://onlinelibrary.wiley.com/doi/10.1046/j.1365-2648.1995.22010048.x/abstract