Constructing Assurance Case using Information From an Issue Tracking System

Khana Chindamaikul†1, Toshinori Takai†1, Hajimu Iida†1

Abstract—We address the problem of constructing an assurance case by presenting an approach to find some issues in issue tracking system that can be information for constructing an assurance case. We take advantage of document retrieval and topic modeling techniques to elicit relevant information which can be used as materials for constructing assurance cases. This paper gives an overview of an approach and reports the result of some preliminary experiment. The results suggest that the proposed approach could be effective in terms of reducing time and cost for constructing an assurance case.
Keywords—assurance case, document retrieval, topic modeling

1. Introduction

Nowadays, many systems tend to be huge. Systems including large software cannot achieve complete safety or security. So, it is very important to provide justified confidence for those systems by an *assurance case* (*AC*). An AC is a documented body of evidence that provides a convincing and valid argument that a specified set of critical claims about a system's properties are adequately justified for a given application in a given environment[1]. However, constructing an AC takes a lot of time, effort and its cost tends to be high. Our work aims to reduce time of understanding whole the related documents and cost of constructing an AC by automatically selecting the relevant documents.

2. Background

2.1 Assurance Case (AC)

An assurance case consists of 3 parts: A *claim*, an *argument* and *evidence*. A claim is a proposition about an attribute or a property of the system, an argument is a description showing how the evidence supports the claim, and evidence is data showing the claim holds, e.g. facts, assumptions, or other ACs.

2.2 Topic modeling

Topic modeling is a technique for automatically extracting semantic topics from a huge collection of text documents[3]. The underlying idea is based on the assumption that each document can be represented by a small number of topics, where each topic is dominated by a small fraction of all possible words. There are many algorithms for topic modeling such as LDA, LSI, PAM, and MU. We use Mallet as a tool for topic modeling[4], which is based on LDA. Figures 1, 2 show the screenshots of the results of Mallet.

#doc	name	topic	proportion	topic	proportion
69	file:/C:/mallet/./txtInput/BugID%5B8069%5D.txt	35	0.08159	13	0.07198
70	file:/C:/mallet/./txtInput/BugID%5B8070%5D.txt	86	0.18335	31	0.06105
71	file:/C:/mallet/./txtInput/BugID%5B8071%5D.txt	13	0.06233	30	0.06233

Figure 1. A set of top *k* topics for each document

Topic	Dirichlet	A set of top n words for each topic
10	0.48544	answer tolerance case float abs max fraction min precision
11	0.48544	modules current feature imo part function module activities imp
12	0.48544	moodle http org check sourceforge net read martin cvs
13	0.48544	login auth password security running case mode petr username

Figure 2. A set of top *n* words for each topic

2.3 Document retrieval

Document retrieval is a technique to elicit documents related to a given query[2]. Many algorithms has been proposed, which are typically based on Boolean models, Vector space models (ex. LSI), and Probabilistic models.

3. Overview of Construction

In this part, we introduce the basic flow of our approach to find some issues in an issue tracking system. First, we use a document retrieval technique to find related issues to given queries. Then, we group issues each of which has the same topic into a same group by topic modeling. Finally, we select relevant issues that can relate to an assurance case from small size of issues in group of interest, since each group is characterized by a finite set of words given by topic modeling.

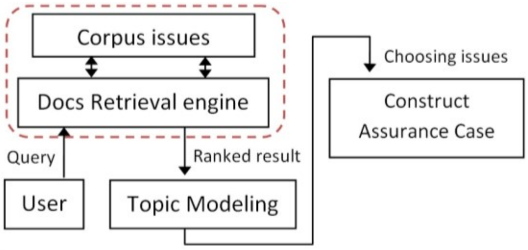

Figure 3. Preparation by using IR and Topic modeling

3.1 Preparation of information for constructing AC

Our approach to prepare information for constructing an AC is shown in Figure 3.

I. Formulating a query: user selects words or sentence that relate to the claim for an AC.
II. Searching document: a set of issues that relates to a query is retrieved (we call the list the *ranked result*) by a document retrieval engine.
III. Applying topic modeling: issues from the ranked result are inferred using a topic modeling tool. Then, issues that are related to the same topic are grouped into the same group. A topic modeling tool can also provide a set of words for each topic. So, user can use some of these words for naming the group. The group that has a name similar to a query is called *group of interest*. Note that some issues relates to many topics. Mallet provides a set of top *k* topics for each issue. Our approach selects the top 2 topics for each inferred issue.
IV. Screening the result: this step, users can easily select issues

†1 Nara Institute of Science and Technology

from a small size of issues in group of interest for constructing ACs.

3.2 Constructing AC

Using information from last preparation step, we can express each of them in argumentation framework (by hand). Then, we can use this information to construct ACs. More details of constructing Assurance Case will be provided in another paper.

4. Design and experiment

We applied our approach for the description of the construction of an AC for a course management system for educational institutes, which is called Moodle, in an issue tracking system, which is called Moodle tracker system. We present 3 different way of construction of the AC and choose the best. The first solution consists in reading all issues and constructing an AC. The second solution is using topic modeling to prepare information for constructing an AC. The third solution is using both document retrieval and topic modeling. For this experiment, we used issues from issue number 8,000 to issue number 8,500 from the corpus issues and the results of those solutions are as follows:

I. The first solution: we read all of the issues. We found that there were only 10 issues that related to safety or security (relevant issues). Those 10 issues will be used for constructing an AC. We used the knowledge from this solution as base knowledge of other solution.

II. The second solution: we used the set of selected issues as an input of Mallet. The number of topics we use as inputs of Mallet is in range from 76 to 115. One of the observations that we had while trying to use less than 76 topics is that words "safety security" did not appear in a set of words for each topic. While when we tried to use more than 115 topics, words "safety security" appeared more than once in a set of words for topics. For example, word “safety” appeared in a set of words for topic1 and it also appeared in a set of words for topic2. From 76 to 115 topics, there was only 1 topic which has words “safety security” in a set of words (Figure 2). To simplify the evaluation we consider the case when words "safety security" appeared only once. We call the topic which has the words “safety” and “security” in a set of words the *safety topic*. After that, we grouped the issues which have top 2 topics related to the safety topic into the same group (see Figure 1, issue8069 and issue8071 were in the same group). This group was called the *safety group*. Then, we analyze the safety group by measuring the precision and recall while we changed the number of topics from 76 to 115. The results are shown in Table 1. Due to space limitations, we present only the top 3 records that had the best precision and recall. And we selected 76 topics as the best input topic number of Mallet for this solution.

III. The third solution: the approach of this solution is shown in Figure 3. For the step in the dotted-line box, we used a search tool built in Moodle Tracker system. First, we used “safety security” as a query. Then, we used the document retrieval method for searching issues in this system. A search engine returns a list of issues relate to the query. In this case, we obtained 44 retrieved issues as the ranked results. After that, we used all issues in the ranked results as input of Mallet. And we used the number of topics ranging from 5 to 26 as an input. The reason why we choose this range is the same as the reason of our second solution. The results of measuring precision and recall of the safety group are shown in Table 1. We selected 5 topics as the best input topic number of mallet for this solution. After that, we constructed an AC from those issues in safety group.

Number of topic	Relevant issues in safety group	Issues in safety group	Total relevant issues	Precision	Recall
Second solution					
76	3	14	10	0.21	0.3
77	2	14	10	0.14	0.2
81	2	10	10	0.2	0.2
Third solution					
5	7	9	10	0.78	0.7
6	3	7	10	0.43	0.3
7	3	5	10	0.6	0.4

Table 1. The result of measure precision and recall

Note: numbers of total relevant issues are from base knowledge.

5. Evaluation

We compare the results of 3 solutions by measuring effort, time, coverage and cost (See Table 2). We assume as the worst case that user will read all issues in the safety group for the solution 2 and 3. Note that, the average time for reading each issue is 4.40 minutes. Effort is measured as the number of issues that a user has to read until the construction of the AC is completed. The coverage is measured by the number of recall.

	Effort	Time(min.)	Coverage	Cost
Solution1	501	2204.4	High	High
Solution2	14	61.6	Low	Low
Solution3	9	39.6	Medium	Low

Table 2. Experimental result from 3 solutions

6. Conclusion and future work

Although the experiment presented here is preliminary phase, the result suggests that the proposed approach, which combines document retrieval and topic modeling can provide useful means for constructing ACs. The result also shows that the approach can be effective because it is possible to reduce the effort, time and cost for constructing. However, some relevant issues may miss by using this approach. So, trade-off between time and coverage of constructing an AC is required for consideration.

We plan to reduce user effort for finding relevant issues in group of interest by using formal concept analysis (FCA). FCA present issues in concept lattice which provide additional structure among issues.

Reference

[1]. Adelard: ASCAD - The Adelard Safety Case Development Manual. 1998.
[2]. Christof Monz. Document retrieval in the context of question answering. ECIR'03, Fabrizio Sebastiani (Ed.). Springer-Verlag, Berlin, Heidelberg. 2003.
[3]. D. Blei. Probabilistic topic models. Communications of the ACM, 2012.
[4]. S. Graham, S. Weingart, I. Milligan. (2012). Getting Started with Topic Modeling and MALLET. From http://programminghistorian.org.

