2022 EECCIS

PROCEEDINGS

11th Electrical Power, Electronics, Communications, Control, and Informatics Seminar (EECCIS 2022)

IEEE Conference Number #54468
IEEE Catalog Number CFP2232Z-ART
ISBN : 978-1-6654-0648-2

Department of Electrical Engineering
Faculty of Engineering
Universitas Brawijaya

IEEE Universitas Brawijaya Student Branch
IEEE Indonesia Section
UTHM Universiti Tun Hussein Onn Malaysia

2022 11th Electrical Power, Electronics, Communications, Controls and Informatics Seminar (EECCIS) | 978-1-6654-0648-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/EECCIS54468.2022.9902934

EECCIS 2022 Title List

A. Electrical Power

2022 11th Electrical Power, Electronics, Communications, Controls and Informatics Seminar (EECCIS) | 978-1-6654-0648-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/EECCIS54468.2022.9902885

B. Electronic Systems

C. Telecommunication

D. Controls System

E. Computer and Informatics

EECCIS 2022 Author Index

2022 11th Electrical Power, Electronics, Communications, Controls and Informatics Seminar (EECCIS) | 978-1-6654-0648-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/EECCIS54468.2022.9902918

EECCIS 2022 Reviewer

Taufik Taufik	California Polytechnic State University (USA)
Ichijo Hodaka	University of Miyazaki (Japan)
Mitsuhiro Yokota	University of Miyazaki (Japan)
Mohd Ashraf Ahmad	Universiti Malaysia Pahang (Malaysia)
Mohammad Faiz Liew Abdullah	Universiti Tun Hussein Onn Malaysia (Malaysia)
Rosli Omar	Universiti Tun Hussein Onn Malaysia (Malaysia)
Jafferi Jamaludin	University of Malaya (Malaysia)
Vikash Bhardwaj	DEWAN VS Group of Institutions (India)
Anggun Isnawati	Institut Teknologi Telkom Purwokerto (Indonesia)
Fauza Khair	Institut Teknologi Telkom Purwokerto (Indonesia)
Suroso Suroso	Jenderal Soedirman University (Indonesia)
Cahya Santosa	National Research and Innovation Agency (Indonesia)
Handrea Tambunan	PLN (Indonesia)
Marwan Marwan	Polytechnic State of Ujung Pandang Makassar (Indonesia)
Brian Pamukti	Telkom University (Indonesia)
Amirullah Amirullah	Universitas Bhayangkara Surabaya (Indonesia)
Dwi Ajiatmo	Universitas Darul Ulum (Indonesia)
Fitri Zulkifli	Universitas Indonesia (Indonesia)
Gunawan Wibisono	Universitas Indonesia (Indonesia)
Teti Zubaidah	Universitas Mataram (Indonesia)
Yudhi Gunardi	Universitas Mercu Buana (Indonesia)
Elta Sonalitha	Universitas Merdeka Malang (Indonesia)
A. Aripriharta	Universitas Negeri Malang (Indonesia)
Ilham Ari Zaeni	Universitas Negeri Malang (Indonesia)
Muladi Muladi	Universitas Negeri Malang (Indonesia)
Vera Suryani	Universitas Telkom (Indonesia)
Gigih Priyandoko	Universitas Widyagama, Malang (Indonesia)
Adharul Muttaqin	Universitas Brawijaya (Indonesia)
Akhmad Zainuri	Universitas Brawijaya (Indonesia)
Angger Abdul Razak	Universitas Brawijaya (Indonesia)
Fakhriy Hario	Universitas Brawijaya (Indonesia)
Lunde Ardhenta	Universitas Brawijaya (Indonesia)
Mochammad Rusli	Universitas Brawijaya (Indonesia)
Muhammad Aziz Muslim	Universitas Brawijaya (Indonesia)
Muhammad Fauzan Edy Purnomo	Universitas Brawijaya (Indonesia)
Onny Setyawati	Universitas Brawijaya (Indonesia)
Panca Mudjirahardjo	Universitas Brawijaya (Indonesia)
Ponco Siwindarto	Universitas Brawijaya (Indonesia)
Primatar Kuswiradyo	Universitas Brawijaya (Indonesia)
Raden Arief Setyawan	Universitas Brawijaya (Indonesia)
Rahmadwati Rahmadwati	Universitas Brawijaya (Indonesia)
Rini Hasanah	Universitas Brawijaya (Indonesia)
Rudy Yuwono	Universitas Brawijaya (Indonesia)
Rusmi Ambarwati	Universitas Brawijaya (Indonesia)
Sapriesty Nainy Sari	Universitas Brawijaya (Indonesia)

2022 11th Electrical Power, Electronics, Communications, Controls and Informatics Seminar (EECCIS) | 978-1-6654-0648-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/EECCIS54468.2022.9902892

Sholeh Pramono	Universitas Brawijaya (Indonesia)
Sigit Kusmaryanto	Universitas Brawijaya (Indonesia)
Tri Kurniawan	Universitas Brawijaya (Indonesia)
Tri Nurwati	Universitas Brawijaya (Indonesia)
Wijono Wijono	Universitas Brawijaya (Indonesia)
Zainul Abidin	Universitas Brawijaya (Indonesia)

Expert Annotation Tools for Labeling Student Capstone Project based on ACM CCS Ontology

Publisher: IEEE

Andre Andre ; Nanik Suciati ; Hadziq Fabroyir All Authors

Abstract: Students enrolled in an undergraduate computer science program must complete a capstone project as their final project. The selection of a capstone topic that is not appr... **View more**

Metadata

Abstract:
Students enrolled in an undergraduate computer science program must complete a capstone project as their final project. The selection of a capstone topic that is not appropriate can cause delays in the completion of this activity. Consequently, a thorough comprehension of the disciplines represented in each capstone title is required as a reference point for students when selecting topics that are appropriate for their academic skills. The popular ontology produced by the Association of Computing Machinery (ACM) can be used to learn about different computer science disciplines. Capstone title labeling on this ontology is applicable in topic recommendation systems that use clustering techniques to group titles into groups of similar topics. This technique requires a manual labeling process carried out by an expert, which has proven to be superior in other studies. However, this process is cumbersome and prone to errors. As a result, we present annotation tools that can assist experts in labeling based on the rules of the ACM Computing Classification System (CCS) ontology. Such tools have never been developed before in the education domain, despite the fact that educational datasets are widely available. In order to evaluate the efficacy, reliability, and accuracy of our system, we employ expert services to make manual annotations with and without the use of tools. The results of the comparison show that the manual labeling process with our tools are able to speed up the annotation process and increase the accuracy of the results

achieved, as well as assist the resolution process of labeling disagreements among experts.

Published in: 2022 11th Electrical Power, Electronics, Communications, Controls and Informatics Seminar (EECCIS)

Date of Conference: 23-25 August 2022

INSPEC Accession Number: 22114255

Date Added to IEEE *Xplore*: 04 October 2022

DOI: 10.1109/EECCIS54468.2022.9902931

ISBN Information:

Publisher: IEEE

Conference Location: Malang, Indonesia

Contents

I. Introduction

Data pre-processing involves transforming raw data into a structured and easy-to-understand form. This procedure creates accurate and reliable datasets which ready to be analyzed in an algorithm, such as machine learning. One of the pre-processing data methods involves expert judgments to assess the data. For instance, in the problem of clustering and classification, especially in big data within education domain with unstructured data, the involvement of experts has been investigated extensively [1]–[4], which produces improved accuracy and reliability. The manual annotation process is challenging-the proclivity includes error labeling, ambiguous decisions, and tremendous complex effort in term of tedious and time-consuming job. However, several studies indicate that expert labeling has superior benefits to non-expert methods. For example, in research [5]–[7] combining manual annotation by expert and crowdsourcing to produce an accurate dataset viewed from the aspect of annotation discrepancy between expert and crowd. While in the other publication [8], [9] developed a multiple expert annotation framework with superior results than the standard. Moreover, besides manual labeling, the expert also contributes to validating automatic labeling by the system [10]–[12].

Sign in to Continue Reading

Authors

Figures

References

Keywords

About IEEE *Xplore* | Contact Us | Help | Accessibility | Terms of Use | Nondiscrimination Policy | Sitemap | Privacy & Opting Out of Cookies

A not-for-profit organization, IEEE is the world's largest technical professional organization dedicated to advancing technology for the benefit of humanity.

Expert Annotation Tools for Labeling Student Capstone Project Based On ACM CCS Ontology

Andre
Department of Informatics
Institut Teknologi Sepuluh Nopember
Universitas Surabaya
Surabaya, Indonesia
andre@staff.ubaya.ac.id

Nanik Suciati
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
nanik@if.its.ac.id

Hadziq Fabroyir
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
hadziq@its.ac.id

***Abstract*— Students enrolled in an undergraduate computer science program must complete a capstone project as their final project. The selection of a capstone topic that is not appropriate can cause delays in the completion of this activity. Consequently, a thorough comprehension of the disciplines represented in each capstone title is required as a reference point for students when selecting topics that are appropriate for their academic skills. The popular ontology produced by the Association of Computing Machinery (ACM) can be used to learn about different computer science disciplines. Capstone title labeling on this ontology is applicable in topic recommendation systems that use clustering techniques to group titles into groups of similar topics. This technique requires a manual labeling process carried out by an expert, which has proven to be superior in other studies. However, this process is cumbersome and prone to errors. As a result, we present annotation tools that can assist experts in labeling based on the rules of the ACM Computing Classification System (CCS) ontology. Such tools have never been developed before in the education domain, despite the fact that educational datasets are widely available. In order to evaluate the efficacy, reliability, and accuracy of our system, we employ expert services to make manual annotations with and without the use of tools. The results of the comparison show that the manual labeling process with our tools are able to speed up the annotation process and increase the accuracy of the results achieved, as well as assist the resolution process of labeling disagreements among experts.**

Keywords—annotation, expert system, ontology, capstone project

I. INTRODUCTION

Data pre-processing involves transforming raw data into a structured and easy-to-understand form. This procedure creates accurate and reliable datasets which ready to be analyzed in an algorithm, such as machine learning. One of the pre-processing data methods involves expert judgments to assess the data. For instance, in the problem of clustering and classification, especially in big data within education domain with unstructured data, the involvement of experts has been investigated extensively [1]–[4], which produces improved accuracy and reliability. The manual annotation process is challenging—the proclivity includes error labeling, ambiguous decisions, and tremendous complex effort in term of tedious and time-consuming job. However, several studies indicate that expert labeling has superior benefits to non-expert methods. For example, in research [5]–[7] combining manual annotation by expert and crowdsourcing to produce an accurate dataset viewed from the aspect of annotation discrepancy between expert and crowd. While in the other publication [8], [9] developed a multiple expert annotation framework with superior results than the standard. Moreover, besides manual labeling, the expert also contributes to validating automatic labeling by the system [10]–[12].

The manual annotation method used by the expert has been shown to be more effective and precise in terms of processing time when the appropriate tools are used to facilitate it. Annotation tools provide a corridor for expert reasoning in labeling or annotating datasets. Furthermore, these tools may provide recommendations and even perform automatic labeling, which makes it easier for experts to complete their tasks efficiently. Several tools also take advantage of the ontology in case studies where the knowledge has been mapped comprehensively and structured in a standardized way. The medical domain is where the majority of the research and development of annotation tools based on ontologies takes place. The annotation method with a crowdsourcing approach to medical data produces a knowledge base for mental health literature, segmentation of cellular biology images, and the retrieval process of electronic medical records [13]–[15]. Annotation tools have also been developed for labeling multimedia data [16], [17] with the help of a previously trained semantic ontology. Annotated multimedia data is beneficial in the retrieval of multimedia data from a database. The findings of this study demonstrate a higher level of accuracy that is better, faster, and more efficient. Despite the fact that there have been various framework or proof of concept annotation tools published, we have discovered that there has only been a limited number of studies that have focused on annotation tools in the educational field.

A. Capstone Project and Ontology

Undergraduate computer science students are required to complete a capstone project as part of their degree. However, incorrect topic selection as well as a lack of cognitive skill in the selected capstone resulted in delays in the completion of the assignment. Among the options is a widely used machine learning clustering techniques for developing a recommendation system. The goal is to group capstones into topic groups which can then be used to assess students' abilities and suitability for certain topic groups. It is necessary to involve experts in the pre-processing of data stage in order to annotate each capstone title for the relevant disciplines in order to use this recommendation method. This annotation procedure makes advantage of established ontology standards in computer science, such as the Computing Classification System developed by the Association of Computing Machinery (ACM CCS).

B. Problem Statement, Gap, and Motivation

The capstone topic recommendation system requires a clustering technique to group similar capstone titles. Several courses support each capstone title, correspondingly the clustering process maps these titles into cluster groups. In this case, the system requires a supervised learning method for the dataset of capstone titles. As a result, in order to build an accurate and efficient dataset, it is necessary for the expert to provide computer science annotations based on the ACM CCS

Ontology on each capstone title. These activities are time-consuming, tedious, and need high concentration and effort. In addition, the current research gap highlights the lack of annotation tools that have ever been developed to label capstone data based on computer science ontology. Based on these problems, the research question is formulated as follows:

RQ 1. How to help experts annotate capstone titles more easily and quickly?

RQ 2. How is the accuracy level of expert labeling results compared to manual labeling without tools?

RQ 3. How can annotation tools highlight inaccurate annotation information and help settle a disagreement between experts?

II. METHODOLOGY

A. *Computer Science Capstone Project*

The capstone project at the undergraduate study level is the final stage determining student graduation. Therefore, students should master the related courses that support a capstone title. Capstone project delaying is a common problem because students lack cognitive incompetence in the fundamental knowledge of a particular capstone title. As a result, students exhibit a low level of esteem and intention to work on the title. A case study at the undergraduate level at the Informatics Engineering University of Surabaya shows that the average student spent more than two semesters working on the capstone project, albeit the standard duration of capstone project completion is one semester. This study involves 301 students who graduated from the 2016-2021 class.

Therefore, students must choose capstone titles and topics that match their abilities and passions. The topic recommendation system can help students to choose a suitable capstone topic. This system utilizes clustering techniques to group capstone topics that maintain similarities. Then the system reads student transcripts to determine the student cluster to provide recommendations for the proper topic. Data pre-processing stages are needed to build a recommendation system based on machine learning clustering. The raw dataset compromises the bulk of the computer science capstone project. The challenge of this topic clustering is that each capstone title has specific computer science knowledge that students must master. It is possible to have cross-domain knowledge within a title. For example, a capstone with the topic of machine learning requires an understanding of statistics; in addition, the student may master mobile development if the machine learning is implemented in a mobile application.

The Association of Computing Machinery: Computing Classification System (ACM CCS) is a computer science ontology system that classifies and identifies computer disciplines that can be accessed on the web https://dl.acm.org/ccs. ACM CCS is a poly-hierarchical ontology applied to modern computer disciplines and reliable for future changes. ACM CCS is used to classify research publications in computer science into 13 main domains. Each domain has a sub-field depth of 4-6 tree levels. This ontology can be used to label the computer science project capstone according to standards, and this process requires expert involvement. The result of this process is a dataset of a capstone project that consists of fundamentals courses, and it requires as the basis of calculation in clustering algorithms applied.

B. *Annotation Tools Framework Design*

Manual annotating is tedious work and prone to mistakes. Therefore, we developed annotation tools to help simplify the work of experts in the dataset labeling process. Fig 1 shows the workflow of the annotation process.

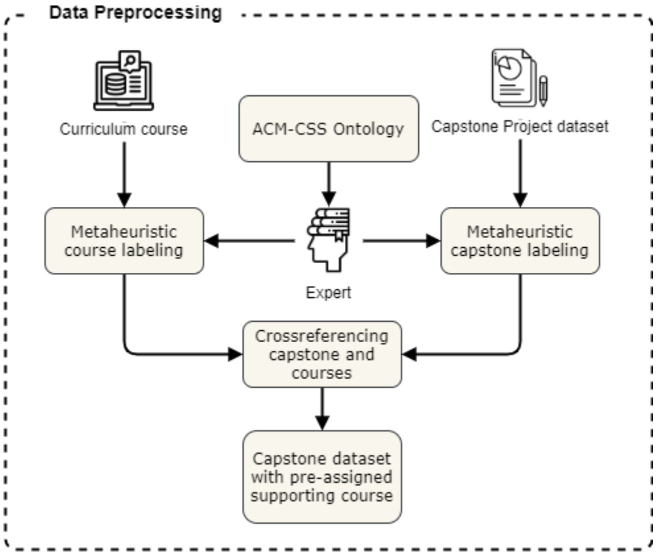

Fig 1. Annotation Tools Workflow Design Process

Four primary processes establish in our workflows that are briefly explained as follows:

1. Understanding of ACM CCS Ontology

Each ontology domain can have a hierarchy of up to a maximum of 4 depth levels. In conjunction with that, the selected ontology keyword should use the deepest hierarchy possible to deliver more accuracy. Consequently, the experts need to familiarize themself with the entire ACM CCS ontology hierarchy. Our annotation tools provide hierarchical tree mapping visualization to help the experts.

2. Metaheuristic course labeling

The final result of the expected dataset is a collection of supporting courses as the fundamental cognitive substantive for each capstone title. For example, in the case study at the Informatics Engineering University of Surabaya, we decided to use 72 courses, of which 23 courses are mandatory, and the rest are elective courses. The step taken by the expert is to label each course based on the ACM CCS ontology manually. Each course covers a broad range of computer science discipline topics. However, not every discipline is prevalent in these subjects. Thus we use coefficients to give the degree of influence of each discipline on the course. The data snippet in the table1 shows the result of the labeling process for each course.

3. Metaheuristic capstone labeling

Furthermore, the expert performs manual labeling of each capstone title according to the rules of the ACM CCS ontology. The selection of ontology keywords uses the most profound level to make the results more accurate. Each capstone title could consist of multi-disciplines of computer science.

4. Cross-referencing capstone and courses

The system uses the cross-reference technique of the ontology keywords of each course against each capstone title. Then the system displays a list of courses that are related or have substantive elements to the capstone title. The expert's task is to select relevant courses to be associated with each capstone title. The results of this final process are capstone titles that have associations with the supporting courses. The clustering algorithm requires this result to generate the expected outcomes.

III. Result And Discussion

This section describes the work plan for implementing the annotation tools in a closed system trial. We involved multiple experts consisting of four experts who served as head of the laboratory responsible for research in each field of computer science. Three experts were assigned to do manual labeling with the help of our annotation tool, while one other expert did manual labeling without the help of tools. To answer the first and second research questions, we measured the input pace of each group and then compared the accuracy of the results. Meanwhile, to answer the third research question, we observed the amount of disagreement data and measured how long the resolution process took.

A. End to End Testing: Study Case Computer Science Undergraduate Capstone Project

A complete and hierarchical-based visual ontology is made available by ACM CCS, which may be accessed on the website located at https://dl.acm.org/ccs. There are a total of 13 core fields of study, and each of these fields has its own hierarchical structure with up to four levels of concentration. The purpose of the ACM CCS is to assist researchers in correctly categorizing their publications so that they may be indexed and retrieved more quickly. Furthermore, our annotation tools provide better visualization of the ontology hierarchy with the help of visual notation in the form of Euler diagrams. As a result, disciplines can be searched and seen more clearly, making it easier for experts to quickly learn and browse each ontology node.

We used 300 capstone titles as dataset testbed from informatics engineering graduates from class 2016 to 2021. Each expert annotated 300 data and then resolved disagreements for specific data. The screenshot display of the annotation tools in Fig 2 shows the ontology keyword input process in a capstone title. Ontology number one shows the disciplines that contributed the most to the title. While ontology number three is assumed to have the lowest contribution to the title. In deciding on the ontology keywords, the expert checks the title, abstract, and even, if necessary, checks the contents of the Capstone project report document. The entered ontology keywords should use the deepest hierarchy of ACM CCS as possible. Furthermore, the system displays the root field of ontology to facilitate the cross-referencing process with the supporting courses that have also been labeled. Finally, the system also displays a table that contains a list of courses associated with these keywords and their coefficient degrees. Information on the list of courses associated with the ontology nodes can help the expert decide on the three supporting courses for the capstone title.

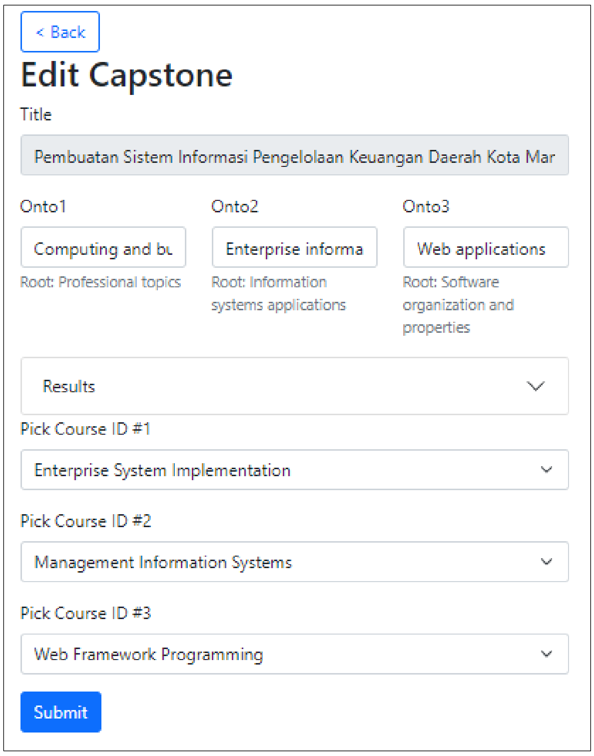

Fig 2. Ontology Input Process on Capstone Title

Using the cross-reference technique, we can investigate the academic substance that emerges between the pre-judgment expert of the capstone title dataset and the courses under consideration. Example: The cross-referencing segment for the capstone title "Development of Decision Supporting Systems Using the Weighted Product Methodologies for Credit Installment of Vehicle Sales" is shown in Table 2. The expert has conducted an assessment and determined that this title requires adequate competence in the fields of *Operation Research, Information System Application,* and *Software notation and tools*. These three fields are ordered from highest to lowest relevance based on the default ACM CCS ontology. Additionally, the mapping process was conducted between comparable courses and these three fields on subjects. The outcome was that it discovered at least nine courses that were relevant to the capstone title categories. To determine the subject relevance between each course and the five aforementioned fields, we employ coefficients ranging from 0 to 1.

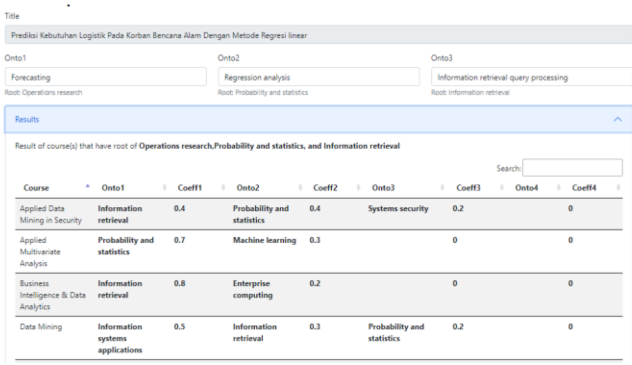

Fig 3. Screenshot of Cross-Reference Result

Table 1. Snippet Of Course Annotation

Courses	Nodes #1	Coeff #1	Nodes #2	Coeff #2	Nodes #3	Coeff #3
Hybrid Mobile Programmi ng	Softwar e notation s and tools	0.7	World Wide Web	0.2	Informa tion systems applicat ions	0.1
Business Geography	Informat ion systems applicati ons	0.8	Informa tion systems applicat ions	0.2	N/A	
Data Mining	Probabil ity and statistics	0.5	Informa tion retrieva l	0.3	Informa tion systems applicat ions	0.2

Likewise, in figure 3 shows how our system displays the results of cross-reference between courses and capstone ontology substance. The system automatically assigns three courses relevant to the title. However, this result can still be overridden by the expert. There is a possibility that the results of each expert's annotations will result in disagreements. In this scenario, disagreement refers to a disparity between the courses that were chosen and the sequence in which they were chosen by the experts. As shown in Figure 4, our system has the capability of highlighting disparities in annotation findings, which can assist in the resolution of arguments amongst experts.

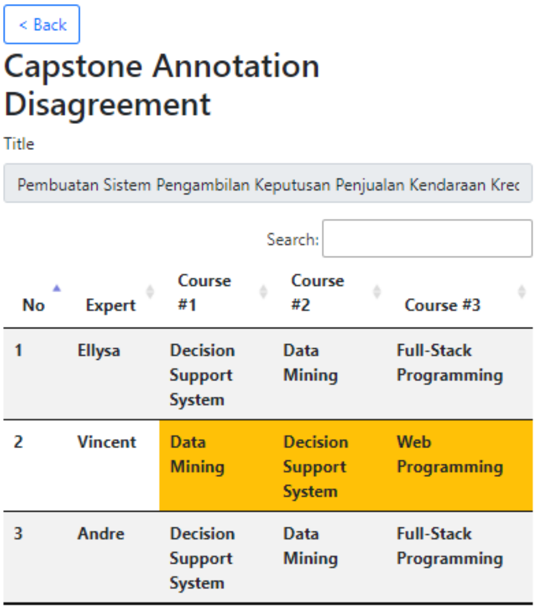

Fig 3. Screenshot of Highlighted Disagreement

B. *Annotation Result Analysis*

The preliminary step is needed before undertaking manual labeling on the capstone title dataset. The chairman of the Department of Informatics Engineering at the University of Surabaya established the 72 courses designated linked to ACM CCS ontology. This annotated course used as ground truth data that need to be consumed in later steps. However, the objectivity of the system that is being designed, the quantity of data, and the knowledge gap between the system designer and the expert all contribute to the quality of the ground truth [18]. A reverse classification framework is utilized as part of an approach in order to make labeling accuracy predictions in the absence of ground truth [19]. To get a higher level of precision and objectivity in the ground truth, one strategy for validating it is to use the assistance of multiple experts in addition to crowdsourcing [20].

Moreover, this technique uses coefficients to represent the degree of considerable importance for each of the relevant ontology nodes for that course. Table 1 illustrates instances of course labeling results. The node with the highest coefficient is considered to influence the course content more than other nodes with a smaller coefficient. This ontology classification becomes the basis for the subsequent capstone labeling process.

In addition, for the purposes of validating the accuracy of our work, we utilized a ground truth dataset that had been verified by the chairman of the Department of Informatics Engineering at the University of Surabaya. We took a sample from 10 different datasets, each of which contained a capstone title with a different topic: three capstone titles with information system subjects, three capstone titles with computer vision topics, and four capstone titles with data science or artificial intelligence issues. We measure the level of correctness of each expert based on the data from the ground truth, and we also measure the speed at which annotations are completed. The results of the measurements are presented in table 2 below. The annotating tools that we built are used by Experts A through C, however Expert D does not utilize any tools. All experts do the same thing, i.e., check the capstone's title, check the abstract, and, if necessary, check the capstone document. On the basis of the results of the examination, the expert provided three complementing courses for the capstone title and supplied scientific annotations in accordance with the norms established by the ACM CCS. The level of accuracy is then determined by comparing the results of the expert's work with the dataset that represents the ground truth. It is the responsibility of the expert to determine the three complementing course and organize them in sequence. In table 2, the correct course (CC) shows the proper number of courses chosen by the expert. Meanwhile, the correct order (CO) indicates if the three complementing courses have been successfully placed in the correct sequence.

Table 2. Result of Annotation Process of Selected Data Sampling

	Expert A			Expert B			Expert C			Expert D		
Capstone	CC	CO	Acc	CC	CO	Acc	CC	CO	Acc	CC	CO	Acc
1	3	1	1	3	1	1	3	1	1	3	1	1
2	2	0	0.5	2	1	0.75	3	0	0.75	3	1	1
3	2	1	0.75	3	1	1	3	1	1	3	1	1
4	3	1	1	3	1	1	3	1	1	3	0	0.75
5	3	0	0.75	3	0	0.75	3	0	0.75	3	0	0.75
6	2	1	0.75	2	0	0.5	3	1	1	3	1	1
7	3	1	1	3	1	1	3	1	1	3	0	0.75
8	3	0	0.75	3	1	1	2	0	0.5	2	0	0.5
9	3	0	0.75	3	1	1	3	1	1	2	1	0.75
10	3	1	1	3	0	0.75	3	1	1	3	0	0.75
	2.7	**0.6**	**0.825**	**2.8**	**0.7**	**0.875**	**2.9**	**0.7**	**0.9**	**2.8**	**0.5**	**0.825**

Accuracy is calculated by adding together CC and CO and dividing by four. In conclusion, each capstone title has three complementing courses that are ordered by their relevance to the title. It can be seen that the use of tools does not significantly increase accuracy when compared to the absence of tools. However, it can be seen in table3, that the annotation processing speed is recorded faster when using tools. On average, using tools took 28 minutes to complete ten datasets compared to 47 minutes taken without using tools.

Table 3. Average Result Summary of Selected Data Sampling

	Course Correctness	Order Corectness	Accuracy	Anot. Speed (min)
Expert A	2.7	0.6	0.825	25
Expert B	2.8	0.7	0.875	32
Expert C	2.9	0.7	0.9	27
Expert D	2.8	0.5	0.825	47

Without using tools, the manual labeling process takes an average of 4.7 minutes per capstone title to label ten capstone titles. If the experts work on labeling for all 300 datasets, it approximately requires 1410 minutes (23.5 work hours). In comparison, the time required for the manual labeling process using tools takes an average of 2.8 minutes per capstone title for approximately 84 minutes. It means that annotation tools have successfully accelerated 60% expert performance.

IV. Conclusions

The manual annotation process is challenging. We propose tools to help experts do this manual annotation to speed up the process and increase its accuracy. The case study used is the ontology labeling process for each capstone project title for undergraduate computer science students. The resulting dataset can be used as fundamental for creating a capstone topic recommendation system that prevents early delays in capstone completion. Based on the measurement of the speed and accuracy level, this tool can help the ontology annotating process quickly when compared to the manual method. However, evidence shows slightly improved annotation accuracy with our tools. Disagreement highlights are considered handy features amongst experts because they can quickly highlight the disagreement data and help resolve the discrepancies. Furthermore, visualization of the ontology hierarchy using a visual notation can help experts learn the structure of computer science ontologies. However, the results of automatic labeling and the cross-reference feature created by the system depend on the outcomes of the initial data setup for the 72 courses because it is utilized as a reference system in making decisions for three relevant courses for a capstone title. It can be difficult to get at ground truth that is both genuine and reliable. When determining the ground truth for course annotations, it is best to involve a large number of stakeholders, particularly the head of the lab and the lecturers for the relevant classes.

This research could go in many different directions. For instance, annotation tools may be built in the education field using a variety of educational datasets to facilitate crowd-assisted labeling. This would enable the labeling process to be completed more quickly, which is especially important for large datasets. Furthermore, the system can be developed to automatically extract relevant keywords in capstone documents using text mining techniques and to map keywords to each course automatically. In conclusion, Our expert annotation tools undoubtedly speed up the undergoing process of semi-automatic dataset labeling. Moreover, the results of the dataset in this study are used for the process of clustering capstone topics to be used as an intelligent recommendation system that can provide guidelines for capstone topics tailored to students' performance records.

Acknowledgements

This research is funded by Ministry of Education, Culture, Research, and Technology, Indonesia under grant of doctoral dissertation research year 2022.

References

[1] H. Zhang, H. Yang, T. Huang, and G. Zhan, "DBNCF: Personalized courses recommendation system based on DBN in MOOC environment," *Proc. - 2017 Int. Symp. Educ. Technol. ISET 2017*, pp. 106–108, 2017, doi: 10.1109/ISET.2017.33.

[2] C. Vo and P. Nguyen, "ST_OS: An Effective Semisupervised Learning Method for Course-Level Early Predictions," *IEEE Trans. Learn. Technol.*, vol. 14, no. 2, pp. 238–256, 2021, doi: 10.1109/TLT.2021.3072995.

[3] S. Agarwal, G. N. Pandey, and M. D. Tiwari, "Data Mining in Education : Data Classification and Decision Tree Approach," vol. 2, no. 2, pp. 140–144, 2012.

[4] M. Durairaj and C. Vijitha, "Educational Data mining for Prediction of Student Performance Using Clustering Algorithms," vol. 5, no. 4, pp. 5987–5991, 2014.

[5] D. Wang, P. Tiwari, M. Shorfuzzaman, and I. Schmitt, "Deep neural learning on weighted datasets utilizing label disagreement from crowdsourcing," *Comput. Networks*, vol. 196, no. June, p. 108227, 2021, doi: 10.1016/j.comnet.2021.108227.

[6] S. Rabiger, G. Gezici, Y. Saygin, and M. Spiliopoulou, "Predicting worker disagreement for more effective crowd labeling," *Proc. - 2018 IEEE 5th Int. Conf. Data Sci. Adv. Anal. DSAA 2018*, no. 1, pp. 179–188, 2019, doi: 10.1109/DSAA.2018.00028.

[7] Q. Hu, Q. He, H. Huang, K. Chiew, and Z. Liu, "A formalized framework for incorporating expert labels in crowdsourcing environment," *J. Intell. Inf. Syst.*, vol. 47, no. 3, pp. 403–425, 2016, doi: 10.1007/s10844-015-0371-6.

[8] H. Valizadegan, Q. Nguyen, and M. Hauskrecht, "Learning classification models from multiple experts," *J. Biomed. Inform.*, vol. 46, no. 6, pp. 1125–1135, 2013, doi: 10.1016/j.jbi.2013.08.007.

[9] Z. Li, Y. Yu, T. Wang, G. Yin, X. Mao, and H. Wang, "HAF: a hybrid annotation framework based on expert knowledge and learning technique," *Sci. China Inf. Sci.*, vol. 65, no. 1, pp. 2021–2023, 2022, doi: 10.1007/s11432-019-9891-5.

[10] C. González-Fernández, A. Fernández-Isabel, I. Martín de Diego, R. R. Fernández, and J. F. J. Viseu Pinheiro, "Experts perception-based system to detect misinformation in health websites," *Pattern Recognit. Lett.*, vol. 152, pp. 333–339, 2021, doi: 10.1016/j.patrec.2021.11.008.

[11] X. Qin, S. He, Z. Zhang, M. Dehghan, and M. Jagersand, "ByLabel : A Boundary Based Semi-Automatic Image Annotation Tool," 2018, doi: 10.1109/WACV.2018.00200.

[12] M. Tamper *et al.*, "AATOS -- A Configurable Tool for Automatic Annotation," in *Language, Data, and Knowledge*, 2017, pp. 276–289.

[13] X. He, H. Zhang, and J. Bian, "User-centered design of a web-based crowdsourcing-integrated semantic text annotation tool for building a mental health knowledge base," *J. Biomed. Inform.*, vol. 110, no. April, p. 103571, 2020, doi: 10.1016/j.jbi.2020.103571.

[14] A. Cocos, T. Qian, C. Callison-Burch, and A. J. Masino, "Crowd control: Effectively utilizing unscreened crowd workers for biomedical data annotation," *J. Biomed. Inform.*, vol. 69, pp. 86–92, 2017, doi: 10.1016/j.jbi.2017.04.003.

[15] S. M. Bafti, C. S. Ang, M. M. Hossain, G. Marcelli, M. Alemany-Fornes, and A. D. Tsaousis, "A crowdsourcing semi-automatic image segmentation platform for cell biology," *Comput. Biol. Med.*, vol. 130, no. September 2020, p. 104204, 2021, doi: 10.1016/j.compbiomed.2020.104204.

[16] R. J. Hendley *et al.*, "CASAM: Collaborative human-machine annotation of multimedia," *Multimed. Tools Appl.*, vol. 70, no. 2, pp. 1277–1308, 2014, doi: 10.1007/s11042-012-1255-1.

[17] R. Lakshmi Tulasi, M. Srinivasa Rao, K. Usha, and R. H.

Goudar, “Ontology- Based Annotation for Semantic Multimedia Retrieval,” *Procedia Comput. Sci.*, vol. 92, pp. 148–154, 2016, doi: 10.1016/j.procs.2016.07.339.

[18] E. J. Favela and K. Yordanova, “Challenges Providing Ground Truth for Pervasive Healthcare Systems,” pp. 100–104, 2019.

[19] V. V Valindria *et al.*, “Reverse Classification Accuracy : Predicting Segmentation Performance in the Absence of Ground Truth,” vol. 36, no. 8, pp. 1597–1606, 2017.

[20] T. Spinde, D. Krieger, M. Plank, and B. Gipp, “Towards A Reliable Ground-Truth For Biased Language Detection,” in *2021 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, 2021, pp. 324–325, doi: 10.1109/JCDL52503.2021.00053.