IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 11, 2019, accepted March 22, 2019, date of publication April 11, 2019, date of current version April 22, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2910145

Automating Articulation: Applying Natural Language Processing to Post-Secondary Credit Transfer

ANDREW HEPPNER, ATISH PAWAR, DANIEL KIVI, AND VIJAY MAGO
DATALab.Science, Department of Computer Science, Lakehead University, Thunder Bay, ON P7B5E1, Canada

Corresponding author: Vijay Mago (vmago@lakeheadu.ca)

This research was funded by the Ontario Council for Articulation and Transfer (ONCAT) and Natural Sciences and Engineering Research Council (NSERC)-Discovery Grant.

ABSTRACT Within the field of post-secondary student mobility, the assessment, and evaluation of transfer credit is a labor-intensive human intelligence task that is subject to time limits and human bias. This paper introduces a semi-automated approach to assessing transfer credit and generating articulation agreements between post-secondary institutions using natural language processing (NLP). The output from the NLP system is tested using a content expert generated an assessment of transfer credit between computer science programs at two separate post-secondary institutions. Initial testing with an unsupervised NLP algorithm, despite good results against standardized measures, assessed the percentage of course overlap as 71% similar to the percentages selected by human content experts. The application of an algorithm based on the Word2Vec model using domain-specific Wikipedia corpus and dependency parsing was applied to compensate for domain specific language and improved the relationship between content experts ratings and NLP output to 86% related overlap.

INDEX TERMS Articulation agreement, natural language processing, semantic similarity, student mobility, transfer credit, word embeddings.

I. INTRODUCTION

The field of student mobility refers to the movement of students between different post secondary institutions and credentials (certificates, diplomas, degrees) throughout their academic and professional career [16], [28]. One of the primary concerns in this field is the transferability of academic credit between institutions and credentials. A continuum of complexity exists with respect to how 'mobile' a credit or degree is within various education contexts [16]. For example, a student completes a Psychology 101 course with a good grade at university A but then moves to university B to be closer to family. As a relatively standard course in many universities, the student will likely get a 'transfer credit' for the Psychology 101 course at university B. Ideally, the student does not have to pay tuition, attend lectures and complete assignments for the same class twice.

When multiple credits and credentials are considered within the context of mobility and transferability across domestic and international post secondary settings, the complexity of student mobility increases. Messacar [24] deftly navigates this complexity by examining transfer credit through an economic lens, similar to others [16]. She presents post-secondary credits as a form of academic currency. The student uses this currency to both gain access to various workplace settings and also to gain access to desired post-secondary education settings. The value of these credits increase or decrease based on the prestige of the institution they were taken at, the relevance of the course content and also subjective human evaluations from admissions personnel and faculty members [20], [28]. Clearly an optimized transfer package that provides credit for relevant and previous learning has the following benefits:

- Saving the student time, ranging from hours per week to years per credential.
- Saving the student tuition money. In 2017 the average cost of full time university tuition per year in Canada is $6571 CAD [1] and the average cost of full time university tuition per year in the US is $9970 USD [13].

The associate editor coordinating the review of this manuscript and approving it for publication was Wenge Rong.

- Decreasing government/taxpayer subsidies related to the support of post-secondary students.
- Returning educated professionals to the workforce as efficiently as possible to benefit society and contribute to the tax base. For instance, state mandated transfer credit pathways have been developed to ease nursing shortages [34].

Despite these obvious benefits, actual data on the average amount and type of transfer credit provided in domestic and international post-secondary settings remains elusive [16], [28] with some reports suggesting that the amount and type of transfer credit presented in the contractual package is lower academic value (currency) than what is expected or, perhaps, fair [12], [28]. Additionally, literature on the processes used to assess and evaluate the transferability/value of academic credit/currency is also elusive [28] although regional examples do exist [20].

The members of this research team were initially brought together by engaging in a year long project to develop a set of standardized transfer credit packages between several post-secondary Computer Science credentials. Our observations in the process were that:

- Assessing the transferability of post-secondary credits between institutions is highly labor intensive and time consuming [20], [28]. Conducting a gap analysis that examines the differences between two related programs is the most intensive task.
- A high degree of variation exists between different faculty members on the value and transferability of the same courses based on their individual perceptions and also, potentially, their personality types.
- The development of transfer pathways is unnecessarily expensive [31] and slow moving through administrative approval structures [28].

As a result of the aforementioned observations, our research team, consisting of Computer Scientists and Transfer Credit professionals, agreed to combine our skill sets to develop a software tool that would make the process of assessing and evaluating transfer credit more efficient, less subjective and less expensive. This paper explores the process, testing and results of an online system that relies on Natural Language Processing (NLP) to automate the process of assessing transfer credits using course specific learning outcomes between post-secondary institutions. All results and testing for this paper are specific to the educational field of Computer Science so that the results of the human process can be compared to the results of the NLP application.

The intention of the project was that stakeholders engaged in transfer credit assessments and transfer credit pathway development could upload their course outlines to an online database whereby a NLP algorithm would be applied to:

- generate a list of overlapping learning outcomes between the courses/programs.
- generate a set of recommended courses to consider for 'pre-defined' transfer credit agreements between high affinity (closely related) programs.

A. MAIN CONTRIBUTIONS

The contributions of this research include:

- The development of a web based system that increases the efficiency of assessing post-secondary transfer credits at multiple institutions
- A new NLP algorithm that provides measures of semantic similarity that can be accurately implemented across multiple specialized domains.

We begin by defining the systems used in the field of post-secondary student mobility (section II) and then provide an overview of NLP applications and technology in the field of Education (section III). The methodology for developing transfer pathways is followed by the web based system architecture and an overview of the applied NLP algorithm (section IV). Testing methodology is then outlined and the results of the web based NLP application are outlined (section V). A discussion (section VI) of our system in terms of both applicability to the field and contributions to the overall field of NLP leads to our conclusions (section VII).

II. MECHANISMS AND PROCESSES IN MOBILITY

The potential of NLP based semantic course comparisons has a relatively large scope. The contribution of an online, automated course content can be applied to many of the mechanisms used to assess and manage transfer credit [5] some of which include:

1) *Course by Course*: Assessing transfer credit on a course by course basis for each individual student, which sometimes involves a significant fee where the results of the assessment are not revealed until the student commits to enroll at the institution [28].
2) *Articulation Agreements*: Developing articulation agreements where a student with a specific credential from a specific insitutions (i.e., Diploma in Nursing from Instition A) gets a specific 'block' of transfer credit towards a closely related program (i.e., Honors Degree in Nursing Science from Institution B).
3) *Regional or Multi-Lateral Pathways*: Developing provincial/ statewide credential pathways where all the holders of a specific credential within that region, from any institution get a specific 'block' of transfer credit (ie. Any Diploma in Nursing completed with the political region gets X number of transfer credits at Institution B towards an Honors Degree in Nursing Science).

The most labor intensive and least reliable of the aforementioned mechanisms is generally course by course assessment on an individual basis. Course to course assessment can be avoided through the development of articulation agreements and pre-determined blocks of credits. These types of agreements have several benefits. They relieve the burden of assessing individual transfer credits during critical points in the academic year, they provide standardized and consistent amounts of credit to each student and, generally, the agreements are transparent to all stakeholders.

A. LEARNING OUTCOMES

To operationalize an NLP application in the context of post-secondary education it is necessary to have a common source of language in which to compare different courses. For the purpose of this research, learning outcomes were selected as the common language and the best fit for this application. As measurable and tangible statements of what a student will learn in a post-secondary course, learning outcomes are used internationally [19], [21].

III. NATURAL LANGUAGE PROCESSING IN EDUCATIONAL CONTEXTS

A. APPLICATIONS

Applications of NLP in the field of education are primarily related to either essay grading [7] or in sequencing course materials and learning processes through language based interaction with students [36].

Van Bruggen *et al.* [36] discuss the application of latent semantic analysis to assessing previous formal learning for students who have applied to an 'open' university. Within this novel context, students are prescribed course work based on their interests, previous learning and desired educational domain. Similar to this project, the authors identify that matching previous learning to assess transferable post-secondary credit and determine mandatory course work is a labor intensive human intelligence task that could be supported by an intelligent automated NLP system. While the authors do not attempt to implement a system, they clearly highlight that the primary challenge of developing an NLP system in an educational context is that it must produce reliable and valid recommendations across many domains.

B. RECENT ADVANCEMENTS IN NLP

Recently developed models that rank semantic similarity are based on neural networks have produced significant improvements in valid and reliable outcomes [6], [11], [18], [27], [35]. One revolutionary model, proposed by Tai *et al.* [35] uses Glove vectors and subsequently Tree-LSTM. Tree-LSTMs generalize the order-sensitive chain-structure of standard LSTMs to tree-structured network topologies. A *siamese adaptation* of LSTM proposed by Mueller and Thyagarajan [27] outperforms the state of the art models. The authors explain the dependency of their model on a simple Manhattan metric. Their method forms a highly structured space whose geometry reflects complex semantic relationships. Performance evaluations for all aforementioned neural network models are trained on SICK dataset [23] and tested on the same dataset.

While methods have improved there are still several challenges in the field of NLP in that the aforementioned models perform poorly when tested on sentences which do not follow the grammar and structure of SICK sentences. The language of learning outcomes can be highly variable and unlikely to follow grammar/structure protocols despite efforts in some PSE sectors to standardize. Additionally, semantic similarity is computed without considering the context of the word according to the sentence it is contained in. The proposed system addresses these crucial issues by compiling a domain-specific corpus to capture the context of words in particular domain and uses an unsupervised algorithm to overcome the problem of lack of training data for semantic similarity in the educational context.

IV. METHODOLOGY

The methodology consists of three overarching sections. *Pathway Development* is discussed to compare and contrast differences in procedural steps used to develop post-secondary transfer pathways between the NLP web based system and more traditional methods of pathway development. The architecture of the web application is discussed and followed by the technical aspects of the *NLP algorithm* to provide context on how semantic similarities are established within the overarching context of NLP. Finally, testing methodology is presented to compare the overlaps between course learning outcomes identified by content experts and the overlaps identified by the NLP based web system.

A. PATHWAY DEVELOPMENT

In the absence of a definitive body of literature informing the technical processes involved in assessing program and course level transfer credit evaluation in North America [28], our project team has applied four years of experience in pathway development to develop the overall framework, see Figure 1. The following outlines the 'system architecture' of 'human based' transfer pathway development process juxtaposed against the automated NLP and web based processes. Based on the regional 'transfer pathway development' funding structure familiar to the authors of this report, the processes listed below typically take one year to complete. Some institutions have required extensions, additional external funding and there are several instances of unsuccessful pathway development projects.

1) COLLECT COURSE OUTLINES

The first step in evaluating transfer credit is to collect and create a database of all the course material being used for assessment. When two institutions are negotiating a formal articulation agreement, all of the course outlines and/or learning outcomes from each program are generally required. Challenges in collecting course information include:

- Extensive email threads where a project lead attempts to collect course outlines or learning outcomes from multiple stakeholders,
- Lack of trust in uploading course outlines to a cloud based storage system (Google Drive)
- Variations in workload, time lines and vacation schedules between institutions and faculty members resulting in long project delays,
- The hesitancy of faculty members to share course information that is considered their intellectual property with a different institution offering a similar credential,

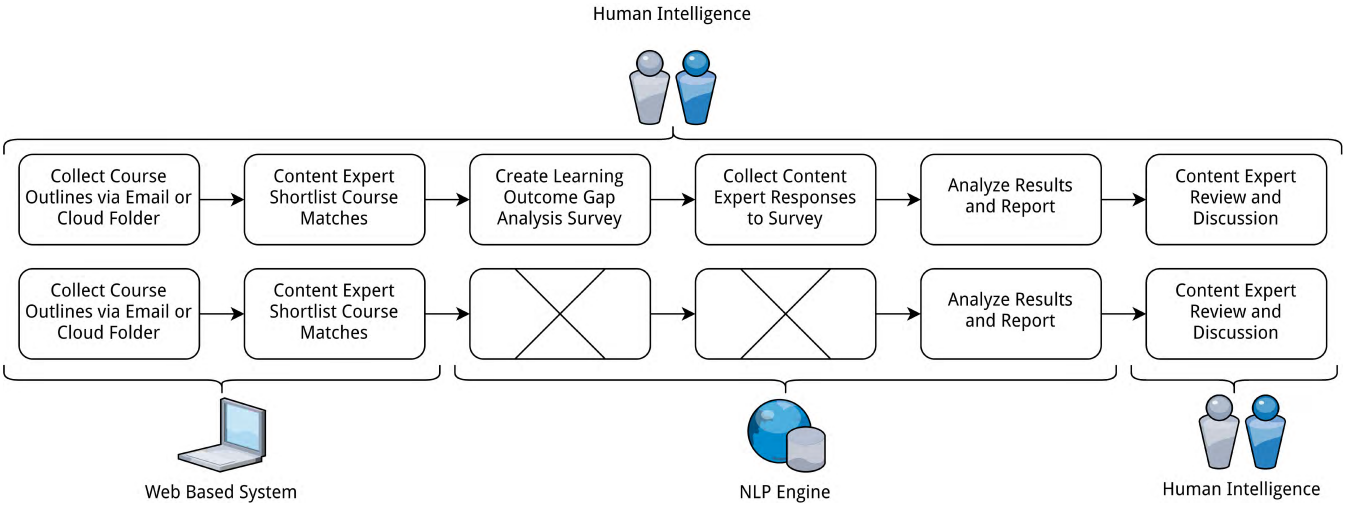

FIGURE 1. Pathway development processes.

Web Based System is designed to optimize the collection of course outlines by allowing faculty members and project team members log in to a password protected, on line system to upload the learning outcomes and, in the future, relevant text associated with a specific course code and department name. This allows faculty members to access and upload course outline information from any location, at any time and also to view and run their own semantic analysis on demand. Additionally, faculty members are only required to upload detailed learning outcomes without necessarily sharing the 'special sauce' that makes their course unique.

2) CONTENT EXPERT SHORTLIST COURSE MATCHES

In our experience, one or more content experts, generally from the lead institution, will do a quick skim of all the course names and course descriptions from both the sending and receiving credential and generate a preliminary set of matching courses. For example, if Institution A has two Data Structures classes and Institution B has one class where the term Data Structures is included in the course description or learning outcomes, that course would be included in the overarching gap analysis. If Institution A has a course on Game Design but Institution B has no content on Game Design, it is likely that this course would be removed from the larger gap analysis.

3) CREATE LEARNING OUTCOME GAP ANALYSIS SURVEY

After establishing a list of potential courses to include in a learning outcome based gap analysis it is beneficial to develop a shared evaluation instrument that produces both quantitative and qualitative data. Typically the evaluation instrument is an online survey that elicits a numerical rating of the overlap between two or more courses and also elicits qualitative feedback from the content expert survey respondent.

The *NLP Engine* is designed in this research uses measures of semantic similarity to generate a preliminary list of matching courses therefore either simplifying the process of negotiating and developing a survey instrument or, ideally, making it unnecessary.

4) COLLECT CONTENT EXPERT RESPONSES TO SURVEY

Depending on the interest level and workload of content experts, it can be challenging to get a high response rate. The evaluation and comparison of learning outcomes between multiple courses is a mentally taxing and time consuming task. A report generated by the web based NLP system summarizes learning outcome overlaps to simplify this task. Additionally this database can fill gaps from incomplete datasets or low response rates to evaluation surveys.

5) ANALYZE RESULTS AND REPORT

Typically, all gap analysis data is collected, analyzed and written up in a report and/or presentation to be shared with relevant stakeholders. The NLP web based system provides a venue for automatically analyzing and generating a report that can be discussed and reviewed by stakeholders early in the process.

6) CONTEXT EXPERT REVIEW AND DISCUSSION

All decisions related to the awarding of transfer credit must be approved by faculty members and then move through all additional administrative layers of approval. It is clear that NLP educational content applications, no matter how accurate, will remain as recommender systems to qualified human education experts [36] but can have a highly valued role in informing and easing the work load associated with assessing transfer credit.

B. NATURAL LANGUAGE PROCESSING ALGORITHM

The proposed gap analysis system uses a previously developed unsupervised semantic similarity algorithm [29]. The method to calculate the semantic similarity between two sentences is divided into two modules:
Pass 1: Maximize the similarity
Pass 2: Bound the similarity

1) PASS 1: MAXIMIZE THE SIMILARITY

This methodology considers the text as a sequence of words and deals with all the words in sentences separately according to their semantic and syntactic structure. The information content of the word is related to the frequency of the meaning of the word in a lexical database or a corpus. A semantic vector is formed for each sentence which contains the weight assigned to each word for every other word from the second sentence in comparison. This step also takes into account the information content of the word, for instance, word frequency from a standard corpus. Semantic similarity is calculated based on two semantic vectors. An order vector is formed for each sentence which considers the syntactic similarity between the sentences. Finally, semantic similarity is calculated based on semantic vectors and order vectors. Pass 1 deals with the three important aspects: Word similarity, Sentence similarity, and Word order similarity [29].

2) PASS 2: BOUND THE SIMILARITY

The first pass of the algorithm returns the maximized similarity (δ) between two sentences. The second pass of the algorithm aims at computing a more robust similarity by reducing the ancillary similarity which causes skewness in results by considering syntactical structure, adjectives and adverbs, and negations in the sentences. Skewness in this context implies the deviation of the similarity(δ) from the similarity in the SICK dataset.

We use *Spacy's dependency parser model* which is the best performing model in the context of this algorithm [29]. The intuitive idea behind this model is to keep track of the syntactical differences by incrementing a global dependency variable. The final similarity(ω) is given by:

$$\omega = \delta - dep\_index \tag{1}$$

where *dep_index* is the dependency index representing the syntactical differences between the sentences [29].

The semantic analysis of any two sentences starts off with the comparison of words in the sentences and thereby determining the semantic similarity between all the words. Hence, the semantic similarity between words is the most crucial aspect when establishing the semantic similarity between sentences. The semantic relations between words are highly domain-specific. In the next sub-section, we describe the use of domain-specific corpus for word similarity.

C. UTILIZING DOMAIN-SPECIFIC KNOWLEDGE FOR WORD SIMILARITY

Learning objectives in Post-Secondary Education contain highly varied words which are rarely used in general English and might have a different meaning across different fields of study. For instance, the computer science domain has hundreds of programming languages such as Python, Java, Lua, etc. The presence of those terms when used in learning objectives make it impossible to use lexical databases such as WordNet [26] because of the limitation of words. Also, it is becoming increasingly difficult to maintain and update the WordNet as it is resource consuming and requires human intervention to redefine or add new relations between words [10]. Instead, we utilize domain-specific knowledge from word vectors formed by word2vec model's *skip-gram* approach using *hierarchical softmax* [25]. The following sub-sections explain the method to compile a corpus using Wikipedia and the word similarity using Word2Vec.

1) BUILDING A DOMAIN SPECIFIC CORPUS

Learning objectives from a course outline contain field specific words. For instance, word 'Python', in the domain of computer science, means 'A programming language' whereas it could mean 'A species of reptiles' in a more general sense. Hence, using a general-purpose corpus is not as accurate a measure of semantic similarity as building a domain-specific corpus and training the model with the corpus. We chose Wikipedia as a source for compiling a corpus [39]. For this research, we focused on a particular domain for the corpus compiled from Wikipedia. We chose 'Computing' as the main sub-category. Wikipedia is divided into multiple sub-categories, and each sub-category can have multiple categories and pages. Figure 3 represents the sub-categories for computing domain.

The *petscan API* gets the Wikipedia structure of a particular category [2]. For this research the total number of Wikipedia articles collected was 160,624. We then applied the Wikipedia python API [14] to retrieve and parse the articles to get the textual content from the article webpage. We store the corpus as a Python file which enables us to compile the corpus to find if there are any non-ascii characters. Filtering such characters is a necessary step before training the model. Every article is stored as a list element in the file for simpler iterations.

2) WORD SIMILARITY

After creating a corpus of *computing* domain, we trained the Word2Vec model with the compiled corpus and used the *gensim*'s implementation of Word2Vec [33].

Through this Wikipedia generated corpus we integrate the previously developed and best-performing unsupervised semantic similarity algorithm with our domain-specific approach.

D. GAP ANALYSIS SYSTEM ARCHITECTURE

The purpose of this application is to give end users from a variety of domains (Admissions/Enrollment, Faculty, Project Coordinators and Upper Administration) the ability to analyze different post-secondary programs and courses and receive a functional analysis of the results.

Implementing NLP gap analysis within a web application presents several challenges. Firstly, the amount of sentence to sentence semantic comparisons that need to be done throughout the lifetime of multiple course to course sets of comparisons is an extremely resource intensive task. It is recommended to take advantage of current cloud infrastructure, the application should be designed to scale from a single small virtual server, to a cluster of large ones.

Since hardware access was limited in this case, the application was designed to work solely on a single cloud instance (7.5GB of memory, 4 virtual CPUs, and 40GB of storage). This server was responsible for the front-end and back-end of the web application, the database to store the results and metadata, running the semantic analysis algorithm, as well any other necessary software. While this scales vertically (providing more computational hardware), all the semantic comparisons will pass through a queuing service before being passed to the semantic analysis service, so that horizontal scaling (providing additional servers to a cluster) can be taken advantage of in the future by placing a load balancer between a centralized queuing server and a cluster of servers dedicated to semantic analysis (which in turn would write to a centralized database server). This was accomplished using a PHP server for the front and back end of the web application, Redis as a queuing service, and MySQL for the database (see Figure 2 for architecture diagram).

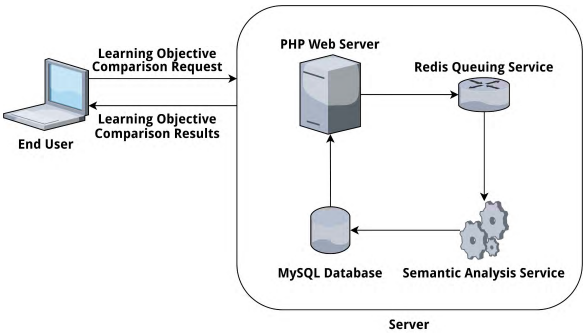

FIGURE 2. Gap analysis system architecture.

FIGURE 3. Wikipedia subcategories for *computing* domain [3] C: Categories, P: Pages.

The user's interactions with the web application are simple using this configuration. After data entry, the rest of the process hands off for the user and completely handled on the server. Users will login to the application and enter all the information they have for each course they wish to include in the comparison. This includes, course name, course number, instructor, and all of the learning objectives for that course. After they have completed entry for each course, they are asked to enter the information for the two programs which will be compared, including the receiving and sending institute and names of the departments. It is at this time that the user will assign each course to its respective program. Once the user submits that information, they can either monitor progress updates given through the application, or they can simply walk away and wait for the application to send them an email once the comparison is complete. In the background, the application saves all of the metadata (course and program information) in the MySQL database, and will start comparing every learning objective from the receiving institute with every learning objective from the sending institute.

Once all of the comparisons have been made, a score is calculated for each combination of receiving and sending course by averaging all of the individual learning objective scores between the two respective courses (Algorithm 1). The learning objective score is simply the highest semantic match made between that learning objective, and any learning objective from the sending course. However, courses that have many semantically similar learning objectives should be preferred over courses that only have one of few similar learning objectives, so the learning objective with the highest semantic score is taken from the course with the most significant semantic scores, if such a course is available.

Input: Receiving and sending courses with learning objectives
Output: Semantic similarity value between 0 and 1
foreach *learning objectives* ***in*** *receiving course* **do**
  foreach *learning objectives* ***in*** *sending courses* **do**
    perform semantic analysis between all receiving learning objectives and sending learning objectives;
    best course ← course with most semantic matches 70% or higher;
    if *best course is found* **then**
      score for learning objective ← highest semantic match that belongs to the best course;
    else
      score for learning objective ← highest overall semantic match;
    end
  end
end
score for course ← average of learning objective scores;

Algorithm 1 Calculating Learning Objective and Course Semantic Rating

E. TESTING METHODOLOGY

To test the validity of the NLP output, data from a historical transfer credit pathway development project in the field of

TABLE 1. Courses and corresponding similarity measures.

Institute 1	Institute 2	Mean Human Similarity	Domain-specific semantic similarity	General purpose semantic similarity algorithm
MATH1271	MATH1033	0.525	0.5804	0.18
CS1431	CS1008 + CS2006	0.775	0.7947	0.57
CS1411	CS1030 + CS2006	0.778	0.7830	0.53
CS2430	CS2125 + CS3025 + CS2068	0.775	0.7836	0.7
CS2477	CS1011	0.787	0.6707	0.66
CS2412	CS2021	0.7	0.6944	0.46
CS3412	CS3002	0.8	0.8133	0.49
CS4453	CS2018	0.5625	0.5852	0.9
CS4411	CS1045 + CS2068 + CS2070 + CS2099	0.3375	0.7846	0.85
CS4478	CS3023	0.65	0.7521	0.86
CS4467	CS3026	0.55	0.7958	0.39

Computer Science was analyzed. We wanted to compare both the semantic similarity of learning outcomes calculated by the NLP system and also the resulting decisions on which courses had the most overlap to the calculations and decisions of content experts. Essentially, would the algorithm produce a list of learning outcome overlaps and recommended course transfer credit overlaps that a content expert would be able to trust? Therefore we attempted to match the NLP process as closely as possible to the human gap analysis process. The historical transfer credit analysis used the following steps:

1) A list of courses from the 'sending institution' were reviewed by one content expert and shortlisted to ensure the overarching project team only analyzed relevant courses. This list was then reviewed by faculty from both institutions and the resulting 'final' list was uploaded into a Google Forms survey.
2) The survey was completed by 9 content experts, 6 faculty from the receiving institution and 3 faculty from the sending institution. Contents experts could: a. view the course outline and learning outcomes from all the pre-selected courses b. rank the course for amount of overlap on a scale of one to ten (one being no overlap and ten being complete overlap), and c. indicate if they recommended giving transfer credit for the course.
3) The results of the online survey were then discussed in person by all stakeholders to reach a final agreement on all the courses which would be given as transfer credit to students from the sending institution.

To mimic this process using the NLP system, we ran our analysis on the same courses that were selected by the content experts.

Linear regressions comparing the overall percentage overlap per course determined by the content experts to the output of the NLP system were conducted.

V. RESULTS

The following section provides an outline of the processing details involved in running the semantic analysis. Course level semantic similarity percentages between the content expert survey, general algorithm and domain specific algorithm are then compared. Finally, a Learning Outcome level semantic analysis output is then explored with an exemplar course.

In our first attempt to run a comparison of the Computer Science course learning outcomes our team decided to run every single course in the programs from both institutions with the hope that it would produce a list of courses from highest matching to lowest matching. The automated learning objective comparison application, using the general purpose semantic similarity algorithm, needs to make all possible course to course comparisons to make a prediction on what two courses will end up matching with each other. There were 78 learning objectives in the set of all learning objectives from the receiving program and 90 learning objectives in the set of all learning objectives from the sending program. Since it is necessary to compare every learning objective in the receiving institute with every learning objective in sending institute, there were 7020 total learning objective comparisons made in this case. This may seem like an insignificant amount, but it is clear that number can quickly increase with additional courses in either of the programs, or when courses have an abnormally large amount of learning objectives. Since we only have access to one server, this process ran synchronously and took 23 hours to complete, but this can be improved with the scaling techniques. Course level results that corresponded with the the courses used in the historical content expert survey were extracted from the overall results for the purposes of comparison.

For the domain specific semantic similarity algorithm, only the required comparisons were made so they could be measured against the content expert survey results. This decision resulting in only 624 comparisons being made, which took 2 hours, but required some manual input.

A. RATINGS OF SEMANTIC SIMILARITY AT A COURSE LEVEL

The combined course level calculations of semantic similarity for both the general purpose and domain specific algorithms were recorded and are listed in Table 3. Percentage overlap is calculated by the converting the similarities onto the scale of 1 to 100 and subsequently considering the difference between the summation of each approach. After running the course outlines and learning outcomes on the general purpose algorithm course level the overlaps in semantic similarity between courses were 71 percent similar to the rankings of content experts. By implementing both dependency parsing

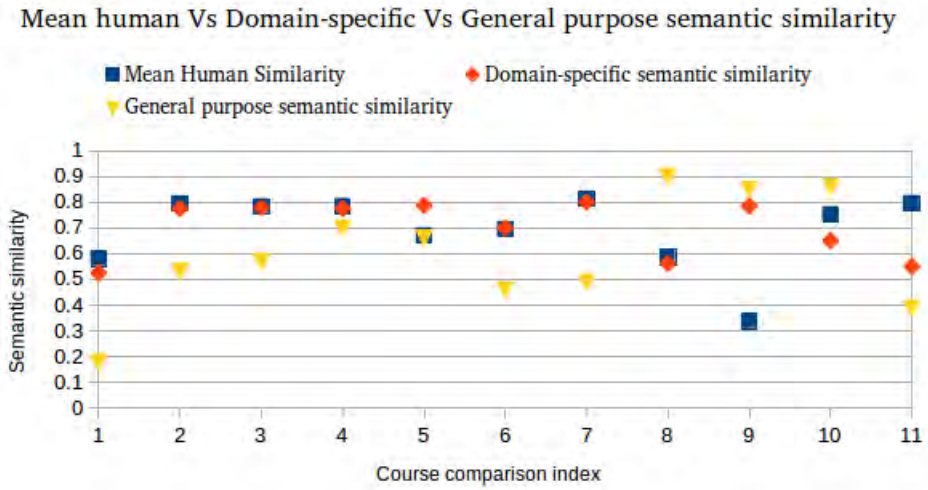

FIGURE 4. Semantic similarity comparison.

and a domain specific Wikipedia based corpus our results in relation to the content expert rankings improved substantially to an overlap of 86 percent. We considered that if the NLP system produced overlaps that more closely matched the content experts rankings, this would indicate increased validity of the NLP system's output.

It is clear in Figure 4 that there is a considerable difference between the general purpose measures of semantic similarity in relation to the domain specific measures. The domain specific algorithm provides a semantic similarity percentage that is closer to the domain experts ranking than the general algorithm in every course in the data set with the exception of Course 5 and Course 9.

When the algorithm is considered as a recommendation system where the courses with the highest semantic overlap are recommended in order, the differences between the general and domain specific algorithm are very influential. Table 4 provides a list of the top five courses that were recommended for consideration credit transfer by each system. Courses in green were approved for credit transfer in the Computer Science and courses in red were not approved for credit transfer. The top three courses recommended by the general purpose algorithm were not considered in the top five courses by content experts. The top three courses recommended by the domain specific algorithm were closer matches where the algorithm gave top ranking to the same course as the content experts. While both 4467 and 4411 are questionable additions, the domain specific algorithm provided more accurate course recommendations than the general purpose algorithm.

B. LEARNING OUTCOME LEVEL

To provide perspective on how the sentence to sentence semantic similarities in learning outcomes differ between the general and domain specific corpus an exemplar course is presented. CS 3412 is selected as it is the most semantically similar course in the domain specific algorithm and the most related course according to the content experts but ranks low on the general algorithm.

Table 3 outlines how the ranking rules 1 arrive at a recommended overall semantic similarity for the course. The first row contains the learning outcomes for CS 3412 from the receiving institution. The second row contains the closest matched learning outcomes as selected by the General NLP algorithm and is followed in row three by the percentage of semantic similarity. The bottom row contains the contrasting learning outcomes as selected by the Domain specific algorithm with respective calculations of semantic similarity above in row four. Clearly, the learning outcome level decisions made by the domain specific algorithm are superior to those made by the general algorithm. While there are still improvements to be made, the difference in overall course percentage is more in line with the ratings of the content experts.

C. SEMANTIC SIMILARITY BETWEEN WORDS

Measures of semantic similarity should provide numbers that are domain specific. Therefore, word comparisons in the computing domain between programming languages such as Java and Python and should have a higher similarity value than they would in a more generalized domain where they would refer to coffee and a species of snake. In Table 2, we can see that, for the first two word-pair comparisons, the semantic similarity from the general purpose corpus is less than the corpus for the computing domain. Similarly, for the word pair 'computing - processing', the general purpose corpora similarities are 0.6369 and 0.3439 for Google News Negatives and English Wikipedia corpora respectively whereas similarity is 0.7118 for the computing domain corpus which makes more sense as computing and processing are closely related to each other in the context of computing. For the words 'ethernet - network', the general purpose cor-

TABLE 2. Comparison of semantic similarity between words.

Word 1	Word 2	Domain-specific word similarity	GoogleNews Negatives 300 similarity	Facebook-fastText [15]using English Wikipedia 2017 16B [38]
java	python	0.9218	0.5626	0.4152
language	python	0.8074	0.5372	0.4052
software	people	0.0406	0.5325	0.3413
computing	semantics	0.7222	0.5986	0.3657
computing	processing	0.7118	0.6369	0.3439
software	application	0.8095	0.6955	0.4284
software	platform	0.7674	0.6777	0.4300
software	tea	0.4869	0.5476	0.2484
java	coffee	0.4098	0.8252	0.4286
database	network	0.7180	0.6710	0.4032
ethernet	network	0.82355	0.7097	0.4442
ethernet	protocol	0.8464	0.6478	0.3937
network	protocol	0.8353	0.6323	0.4530
drivers	firmware	0.8586	0.5794	0.3714
compiler	debugger	0.8659	0.8052	0.5354
windows	linux	0.6868	0.8089	0.5918
windows	door	0.5977	0.7273	0.4567
windows	.exe	0.6475	Word *.exe* not in vocabulary	0.5373
augmented	simulated	0.7742	0.5840	0.4899
google	baidu	0.8655	Word *baidu* not in vocabulary	0.5498
learner	classifier	0.8595	0.6031	0.4458
abstract	namespace	0.7671	0.6200	0.2939

TABLE 3. Semantic similarity between learning outcomes.

CS3412 Database Management	Understand the basic concepts of relational database design and development	Learn the ER diagram methodology of database design	Understand concepts of relational algebra and relational calculus	Learn SQL programming at the basic level, intermediate level and advanced levels	Learn functional dependency and normalization theory	Being able to develop stand-alone/Web based database application using relational database management systems
General Algorithm Top Learning Outcome Selection	design database objects using SQL procedural language	design database objects using SQL procedural language	design database objects using SQL procedural language	create logical and physical 3rd normal form entity models	design database objects using SQL procedural language	define and maintain a database using SQL and a DBMS interface
General Semantic Similarity	0.85	0.88	0.15	0.08	0.09	0.86
Domain Semantic Similarity	0.8058	0.8627	0.8674	0.7933	0.7981	0.7528
Domain Algorithm Top Learning Outcome Selection	create logical and physical 3rd normal form entity models	define and implement constraints to ensure data integrity	define and implement constraints to ensure data integrity	define and maintain a database using SQL and a DBMS interface	design database objects using SQL procedural language	define and maintain a database using SQL and a DBMS interface

pus similarities are 0.7097 and 0.4442, and the computing corpus similarity is 0.8235. We can observe the similar pattern for other word pairs such as 'abstract - namespace', 'learner - classifier'. Hence using general purpose corpus for the computing domain would be inaccurate as ethernet and network are closely related to each other in the context of computing. Also, some of the essential words in the computing domain are not found in the Google News Negatives 300 general purpose corpus, e.g. '.exe' and 'baidu' and even if they are present in the English Wikipedia corpus, the results are poor. Overall from Table 2, we can observe that the semantic similarities between words in the context of computing are better when our domain-specific corpus is used.

TABLE 4. Course recommendations ordered by semantic ranking.

Rank	Domain	Human	General
1	CS 3412	CS 3412	CS 4453
2	CS 4467	CS 2477	CS 4478
3	CS 1431	CS 1411	CS 4411
4	CS 4411	CS1431	CS 2430
5	CS 2430	CS 2430	CS 2477

VI. DISCUSSION AND FUTURE WORKS

By simplifying and expediting the process of developing transfer credit agreements, our overall intention is to save students time and money spent on taking repeat course content. In our original transfer pathway project the process of collecting course outlines, refining learning outcomes and developing a survey instrument that faculty members from two different post-secondary institutions agreed on took approximately five months of work. The completion of the survey and resulting reports and analysis took around two months and significant paid and in-kind efforts on behalf of the entire team. In preparing this paper one of the faculty reviewing it stated that the authors should make sure to, ''add time consuming and painful experiences that you (Project Manager) put faculty members to go through.'' Considering that eight months of work was required to develop a report which faculty members could discuss together in person, the generation of a list of recommendations within the aforementioned processing times is a significant achievement.

The primary challenge in most applications of NLP is to provide measures of semantic similarity that are domain and context specific [4], [8], [9], [17], [22], [30], [32], [37]. This is especially true of any application that is informing faculty and content experts in a post-secondary education setting that specializes in obscure, domain specific language. Specific to the awarding of transfer credit, the faculty members have a high level of authority with respect to their course content and therefore their decisions and opinions are of utmost important in the development of this software. If an NLP system provides a list of learning outcome overlap percentages that a faculty member does not find trustworthy, than it is unlikely that this application will benefit stakeholders. Additionally, this is why content experts opinions are considered as the gold standard for this research.

In relation to the differences between the content expert ratings and domain specific ratings shown in Figure 4 for course 5 and course 9, our perception is that the difference for course 5 is within an acceptable range. We suspect that the reason both algorithms assigned a high percentage of semantic similarity for course 9 is likely due to the high amount of courses it was compared to. We observed throughout the development process that the percentage ranking of semantic similarity is likely to increase when more language is added. In future tests of this application, the percentage ranking of semantic similarity may benefit from being scaled in relation to the amount of courses being compared.

Currently our recommender system still relies on human intelligence to determine the most closely related courses for comparison before processing. Future work could focus on increasing domain specific semantic similarity rankings to the extent that entire programs could be compared to generate regional transfer agreements for multiple institutions. Additionally, the incorporation of recognized taxonomies of learning (Bloom's/Biggs) may assist in ranking and assessing courses for transfer credit.

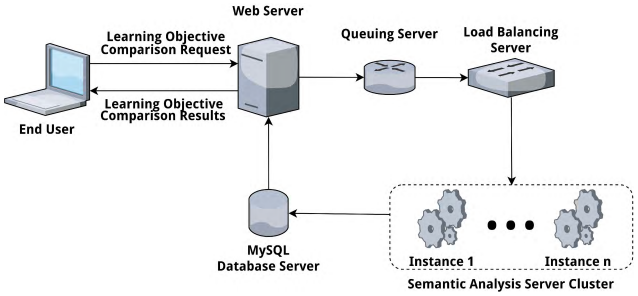

FIGURE 5. Suggested gap analysis system architecture.

A. THE HUMAN FACTOR

This research is based on the assumption that the faculty members are the gold standard in determining the differences courses however there are several human factors that could be considering in future transfer related applications:

1) Volume of Data: If a two year program is compared to four year program, where every year has ten courses with five learning outcomes; a content expert is unlikely to find time to review 300 learning outcomes. Therefore it is possible that content experts under time constraints might rely on heuristics to process the relationships between courses. Our web based system allows for a systematic review of every single learning outcome and may be able to elicit relationships and transfer credit that a heuristic review could miss.
2) Transfer Culture: Depending on the regional context, different types of post-secondary credentials can exist in perceived hierarchies. This can sometimes serve to distract content experts from the specifics of the learning outcomes into a mind set where one course from a prestigious credential is worth two or more from another credential. In this instance, our web system provides an unbiased overview of the learning outcomes.
3) Quality of Learning Outcomes: This system relies heavily on the text of learning outcomes. If the learning outcomes are poorly written or do not accurately represent the actual outcomes of a course, this application will provide invalid results.

B. PROCESSING IMPROVEMENTS

If implementing a similar automated gap analysis tool in the future, and computing resources are available to do so, using scaling techniques is strongly recommended. The high amount of comparisons needed to compare all the courses from two institutes with each other, as well as the computational power needed to make a semantic comparison, means this application can benefit heavily from both horizontal and vertical scaling. Doing so is a simple change from our current implementation (as seen in Figure 2) by isolating all the services onto their own server, creating a cluster of servers dedicated to semantic analysis, and inserting a load balancer between the queuing server and the semantic analysis cluster (Figure 5).

VII. CONCLUSIONS

The benefits of applying a web based system that relies on an unsupervised, customizable NLP algorithm to the process of assessing transfer credit and developing transfer agreements is significant. We would estimate that this recommendation system could potentially shorten the typical 12 month process of developing a transfer agreement by 6 months. By providing a report and analysis of semantic similarities at a course and learning outcomes level, this application can spark discussions among content experts that previously took many hours of group meetings, data collection and analysis to prepare for.

Benefits of the algorithm and web application include:

- Capacity for faculty members to upload and share course outlines and/or learning outcomes from remote locations at any time;
- Relatively high-speed comparisons of learning outcomes and course match sorting that is based on a systematic review of every word and sentence in the entire data set. This is a task by which many faculty members and upper administrators do not have the time to do and may result in insights overlooked by the project team.
- A cloud based system and database for storing credit transfer decisions and analysis for future reference.
- A learning outcome semantic comparison database which can assist in informing necessary approval documents for new transfer credentials.

REFERENCES

[1] *The Daily—Tuition Feeds for Degree Programs, 2017/2018*. Accessed: Sep. 18, 2018. [Online]. Available: https://www150.statcan.gc.ca/n1/daily-quotidien/170906/dq170906b-eng.htm

[2] Wikimedia Foundation. *Petscan API*. Accessed: Jul. 5, 2018. [Online]. Available: https://petscan.wmflabs.org/

[3] Wikimedia Foundation. *Wikipedia Categories*. Accessed: Sep. 18, 2018. [Online]. Available: https://en.wikipedia.org/wiki/Category:Computing

[4] V. Abhishek and K. Hosanagar, "Keyword generation for search engine advertising using semantic similarity between terms," in *Proc. 9th Int. Conf. Electron. Commerce*, 2007, pp. 89–94.

[5] T. H. Bers, "Deciphering articulation and state/system policies and agreements," *New Directions Higher Educ.*, vol. 2013, no. 162, pp. 17–26, 2013.

[6] J. Bjerva, J. Bos, R. V. der Goot, and M. Nissim, "The meaning factory: Formal semantics for recognizing textual entailment and determining semantic similarity," in *Proc. 8th Int. Workshop Semantic Eval. (SemEval)*, 2014, pp. 642–646.

[7] J. Burstein, "Opportunities for natural language processing research in education," in *Proc. Int. Conf. Intell. Text Process. Comput. Linguistics*. Berlin, Germany: Springer, 2009, pp. 6–27.

[8] G. Erkan and D. R. Radev, "Lexrank: Graph-based lexical centrality as salience in text summarization," *J. Artif. Intell. Res.*, vol. 22, no. 1, pp. 457–479, 2004.

[9] A. Freitas, J. Oliveira, S. O'Riain, E. Curry, and J. C. P. da Silva. "Querying linked data using semantic relatedness: A vocabulary independent approach," in *Natural Language Processing and Information Systems*, vol. 8. Berlin, Germany: Springer, 2011, pp. 40–51.

[10] G. A. Miller. *WordNet Online*. Accessed: Sep. 18, 2018. [Online]. Available: https://wordnet.princeton.edu/

[11] Z. He, S. Gao, L. Xiao, D. Liu, H. He, and D. Barber, "Wider and deeper, cheaper and faster: Tensorized lstms for sequence learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 1–11.

[12] N. Heath, "Student mobility in canada across canadian jurisdictions," Pan-Can. Consortium Admissions Transf., Windsor, ON, Canada, Tech. Rep. 20120418, 2012.

[13] M. Jennifer, B. Sandy, P. Matea, and W. Meredith, "Trends in college pricing 2017," The College Board, New York, NY, USA, Tech. Rep. 20120418, 2017.

[14] Jonathan Goldsmithl. (2018). *Python Wikipedia API*. [Online]. Available: https://pypi.org/project/wikipedia/

[15] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov. (2016). "Bag of tricks for efficient text classification." [Online]. Available: https://arxiv.org/abs/1607.01759

[16] S. Junor and A. Usher, "Student mobility & credit transfer: A national and global survey," Educ. Policy Inst., Bethesda, MD, USA, Tech. Rep. ED529950, 2008.

[17] Y. Ko, J. Park, and J. Seo, "Improving text categorization using the importance of sentences," *Inf. Process. Manage.*, vol. 40, no. 1, pp. 65–79, 2004.

[18] A. Lai and J. Hockenmaier, "Illinois-LH: A denotational and distributional approach to semantics," in *Proc. 8th Int. Workshop Semantic Eval. (SemEval 2014)*, 2014, pp. 329–334.

[19] T. Leney, J. Gordon, and S. Adam, *The Shift to Learning Outcomes: Policies and Practices in Europe*. Thessalonika, Greece: Cedefop, 2008.

[20] M. C. Lennon, A. Brijmohan, E. Lavigne, J. Yang, M. Gavin, and L. M. Wheelahan, "Ontario student mobility: Carving paths of desire," Toronto: Centre Study Can. Int. Higher Educ., OISE, Univ. Toronto, Toronto, ON, USA, Tech. Rep. 2015-05, 2016.

[21] M. C. Lennon *et al.*, *Tuning: Identifying and Measuring Sector-Based Learning Outcomes in Postsecondary Education*. Toronto, ON, Canada: Higher Education Quality Council of Ontario, 2014.

[22] P. W. Lord, R. D. Stevens, A. Brass, and C. A. Goble, "Investigating semantic similarity measures across the Gene Ontology: The relationship between sequence and annotation," *Bioinformatics*, vol. 19, no. 10, pp. 1275–1283, 2003.

[23] M. Marelli, S. Menini, M. Baroni, L. Bentivogli, R. Bernardi, and Roberto. Zamparelli, "A sick cure for the evaluation of compositional distributional semantic models," in *Proc. LREC*, 2014, pp. 216–223.

[24] C. Messacar, "Transfer literacy: Assessing informational symmetries and asymmetries," Ph.D. dissertation, Leadership, Higher Adult Educ., Univ. Toronto, Toronto, ON, Canada, 2015.

[25] T. Mikolov, K. Chen, G. Corrado, and J. Dean. (2013). "Efficient estimation of word representations in vector space." [Online]. Available: https://arxiv.org/abs/1301.3781

[26] G. A. Miller, "WordNet: A lexical database for English," *Commun. ACM*, vol. 38, no. 11, pp. 39–41, 1995.

[27] J. Mueller and A. Thyagarajan, "Siamese recurrent architectures for learning sentence similarity," in *Proc. AAAI*, 2016, pp. 2786–2792.

[28] A. P. Ott and B. S. Cooper, "Transfer credit evaluations: How they are produced, why it matters, and how to serve students better," *College Univ.*, vol. 89, no. 4, p. 14, 2014.

[29] A. Pawar and V. Mago. (2018). "Calculating the similarity between words and sentences using a lexical database and corpus statistics." [Online]. Available: https://arxiv.org/abs/1802.05667

[30] T. Pedersen, S. V. Pakhomov, S. Patwardhan, and C. G. Chute, "Measures of semantic similarity and relatedness in the biomedical domain," *J. Biomed. Inform.*, vol. 40, no. 3, pp. 288–299, 2007.

[31] A. Penner and T. Howieson, "Measuring the cost of credit transfer at small colleges," Ontario Council Articulation Transf., Toronto, ON, Canada, Tech. Rep. 2015-02, 2016.

[32] C. Pesquita, D. Faria, A. O. Falcão, P. Lord, and F. M. Couto, "Semantic similarity in biomedical ontologies," *PLoS Comput. Biol.*, vol. 5, no. 7, 2009, Art. no. e1000443.

[33] R. Řehůrek and P. Sojka, "Software framework for topic modelling with large corpora," in *Proc. LREC Workshop New Challenges NLP Frameworks*, 2010, pp. 45–50.

[34] J. Spencer, "Increasing RN-BSN enrollments: Facilitating articulation through curriculum reform," *J. Continuing Educ. Nursing*, vol. 39, no. 7, pp. 307–313, 2008.

[35] K. S. Tai, R. Socher, and C. D. Manning. (2015). "Improved semantic representations from tree-structured long short-term memory networks." [Online]. Available: https://arxiv.org/abs/1503.00075?context=cs

[36] J. V. Bruggen *et al.*, "Latent semantic analysis as a tool for learner positioning in learning networks for lifelong learning," *Brit. J. Educ. Technol.*, vol. 35, no. 6, pp. 729–738, 2004.

[37] G. Varelas, E. Voutsakis, P. Raftopoulou, E. G. M. Petrakis, and E. E. Milios, "Semantic similarity methods in wordNet and their application to information retrieval on the Web," in *Proc. 7th Annu. ACM Int. Workshop Web Inf. Data Manage.*, 2005, pp. 10–16.

[38] Wikimedia. (2018). *Wikipedia Data Dumps*. [Online]. Available: https://dumps.wikimedia.org/archive/

[39] T. Zesch, I. Gurevych, and M. Mühlhäuser, "Analyzing and accessing Wikipedia as a lexical semantic resource," *Data Struct. Linguistic Resour. Appl.*, 2007, Art. no. 197205.

ANDREW HEPPNER is currently a Pathways Coordinator of Lakehead University and has developed many post-secondary transfer credit agreements between Ontario institutions. He has completed an extensive two year mixed methods research study into the experiences and satisfaction of post-secondary transfer students in 2016. He has successfully applied and received funding to develop an automated NLP-based system to assess transfer credit, in collaboration with V. Mago. In addition to his work in student mobility, he is currently a Lecturer in the field of Leadership with the Group Dynamics and Recreation Therapy, Lakehead University. He has been authored and coauthored several reports in the area of student mobility, since 2014.

ATISH PAWAR received the B.E. degree (Hons.) in computer science and engineering from the Walchand Institute of Technology, India, in 2014, and the master's degree (Hons.) in computer science with Lakehead University, in 2018. He was with Infosys Technologies, from 2014 to 2016. He served as a Research Assistant with the DataLaboratory, Lakehead University. He is currently a Software Developer at Intrideo. His research interests include machine learning and natural language processing.

Daniel Kivi is currently pursuing the B.Sc. degree (Hons.) in computer science with Lakehead University, Thunder Bay, ON, Canada. He is currently assisting the Natural Language Processing (NLP) research, for developing technologies to run large NLP tasks and provide interactive results from those tasks. He brings years of experience from his time as a freelance web developer and providing technological services to many communities and events around Thunder Bay. He has taken up several leadership roles at Lakehead University, being the recipient of the Lakehead Luminary Award for excellence in leadership in 2018, and the President of the Lakehead University Computer Science Society and part of the executive team for the Lakehead University Math Club.

VIJAY MAGO received the Ph.D. degree in computer science from Panjab University, India, in 2010. In 2011, he joined the Modelling of Complex Social Systems Program at the IRMACS Centre of Simon Fraser University. He is currently an Associate Professor with the Department of Computer Science, Lakehead University, Thunder Bay, ON, Canada, where he teaches and conducts research in areas, including big data analytics, machine learning, natural language processing, artificial intelligence, medical decision making, and Bayesian intelligence. He has served on the program committees of many international conferences and workshops. In 2017, he joined Technical Investment Strategy Advisory Committee Meeting for Compute Ontario. He has published extensively (more than 50 peer reviewed articles) on new methodologies based on soft computing and artificial intelligence techniques to tackle complex systemic problems, such as homelessness, obesity, and crime. He currently serves as an Associate Editor for IEEE Access and *BMC Medical Informatics and Decision Making* and as a Co-Editor for the *Journal of Intelligent Systems*.