# COMPUTATIONAL MODELS OF NOVELTY BASED ON TOPIC MODELING

by

Maryam Mohseni

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Computing and Information Systems

Charlotte

2023

Approved by:

Dr. Mary Lou Maher

Dr. Celine Latulipe

Dr. Heather Lipford

Dr. Jennifer Weller

# ABSTRACT

MARYAM MOHSENI. Computational Models of Novelty based on Topic Modeling. (Under the direction of DR. MARY LOU MAHER)

Novelty modeling in unstructured text data has been one of the research concentrations within the Natural Language Processing (NLP) community over the past few years. Effective novelty models can play a key role in providing relevant and interesting content to the users which is the central goal in many applications including educational recommender systems. Computational models of novelty provide formal representations for evaluating and generating creative artifacts in creativity and cognition research. Advances in Natural Language Processing provide new approaches to evaluating computational novelty in unstructured text to be applied in multiple cross-disciplinary research areas including Artificial Intelligence, Education, and Human-Computer Interaction.

The problem of novelty measurement in the domain of text has been investigated from different perspectives for different types of textual data. A less examined approach for modeling novelty in unstructured text documents is using Topic Models as the data representation method for gauging computational novelty in research publications. Topic Modeling is a machine learning approach that derives the main themes of a corpus of text documents and represents how they relate. Representing documents with Topic Models has properties that facilitate using various methods for modeling novelty in research publications and also learning materials to be recommended in educational recommender systems.

In this dissertation, we first define a framework for characterizing computational models of novelty that is independent of the type of data in the items. This framework enables an exploration and comparison of existing approaches to computational novelty. We then describe and explore an educational recommender system called

Pique that applies computational models of novelty to encourage curiosity and self-directed learning by presenting a sequence of learning materials that are both novel and personalized to learners' interests. We demonstrate how our computational novelty framework can be applied as the AI component of (educational) recommender systems like Pique, and the usefulness of applying computational models of novelty in educational recommender systems to encourage students' curiosity for expanding their knowledge. We report the student experiences with Pique in four university courses that applied Pique. Based on a qualitative analysis, the students' experience with Pique encouraged their curiosity and led them to unexpected topics in their projects. We then develop two computational approaches to measuring novelty in research publications using Topic Modeling results and demonstrate these models on a database of research publication abstracts from the ACM CHI Symposia. We analyze and describe how the two novelty models differ in the results and interpretation of novelty. Finally, we compare the computational models of novelty based on Topic Models with human perception of novelty by running a study and recruiting experts in the domain of our dataset (HCI) and report on the results. The qualitative analysis of the results suggested that the novelty model based on topic co-occurrence is slightly closer to human perception of novelty compared to the novelty model based on topics similarity. We also found that the criteria for evaluating novelty of a research publication in humans may not be a complete match with the computational models suggesting these two could complement each other.

# DEDICATION

To my mother, Farrokh Ranjbar, who was my best friend, supported me faithfully in my life, and always wished to see my success.

# ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Mary Lou Maher for all her support and mentorship during my PhD study at UNC Charlotte. I am grateful to her patience, knowledge, and motivation guiding me in my academic work. I would also like to thank my dissertation committee members, Dr. Celine Latulipe, Dr. Heather Lipford, and Dr. Jennifer Weller for their supportive and encouraging feedback.

I am grateful to UNC Charlotte Graduate School for providing me with the Graduate Assistant Support Plan funding my PhD study. I would also like to appreciate Dr. Mohamed Shehab, PhD Graduate Program Director, and Sandra Krause for their honest support and kindness during my study at UNC Charlotte.

Finally, I would like to thank specifically my parents for supporting me in my life and always encouraging me to pursue higher education.

# TABLE OF CONTENTS

LIST OF TABLES

# LIST OF FIGURES

# CHAPTER 1: INTRODUCTION

## 1.1 Research Motivation

In recent years the explosion of information has provided a new window of opportunities for extracting and applying useful knowledge from different sources to improve the performance of intelligent systems. These capabilities have led to significant advances in many domains like education and scientific studies in which novelty and creativity are of importance and sought. Novelty is considered an essential characteristic of creativity, along with value and surprise, or unexpectedness [2–4]. Novelty describes how an item differs from those that have come before it, presenting something that did not exist before in that particular form/arrangement [5]. With the increasing use of computational systems to describe and record the products of research, design and creativity, there is an opportunity to extend the evaluation of novelty beyond human assessment [6] and develop computational models of novelty [2, 7].

Evaluating the novelty in text documents plays an important role in domains like education and research because of its role in distinguishing a new document from those that came before it. This is a challenge for human experts as the number of documents increases beyond the ability of any one person to experience and compare [8]. We now have large repositories of documents accessible through web search engines and recommender systems, and curated in databases for learning systems. There is an opportunity to extend our learning systems with computational models of novelty to encourage curiosity and creativity [9]. Observing novel objects can lead to creativity, and being creative is the basis for producing more novel documents [2]. Humans feel a desire to learn more about novel and unexpected objects [10]. Computational models of novelty as the basis for surprising recommendations in recommender systems increase the user's curiosity to explore beyond what they know [11].

The main contributions of this dissertation are: 1) Defining a framework through which novelty models and existing solutions can be categorized, explored, and compared. 2) describing and exploring an educational recommender system called Pique demonstrating the usefulness of applying computational models of novelty in educational recommender systems 3) Describing two computational approaches to measuring novelty using Topic Modeling results and demonstrating these models on a database of research publications abstracts. 4) Finding/demonstrating that different novelty models identify different items as novel showing that novelty has different meanings. 5)comparing computational models of novelty with human perception of novelty by running a survey study recruiting experts in the domain of our dataset.

Different studies investigate modeling novelty in unstructured text data each with a different source data type, representation method, and novelty measure. We define a framework for exploring and categorizing novelty models and existing solutions, independent of the domain and data type. Our framework consists of three main components (shown in Figure 2.1) each of which reflects one major aspect in the analysis of novelty: 1)type of source data, 2)representation method, and 3)novelty model. Each of these aspects has a considerable effect on the performance of intelligent systems for evaluating novelty (see chapter 2 for more details on the framework and its components). Source data (including news, recipes, scientific papers, etc.) needs to be represented in some way so that novelty models can be built on them. Bag of words, word embedding, and Topic Modeling are some methods for representing the text data [12–14]. Regarding novelty models, various approaches to measure novelty of text data are proposed in the literature including similarity based measurements, probabilistic models, and information theory based approaches [11, 13, 15]. However, few investigated the novelty of research publications as one of the unstructured text data types. Also, few have considered Topic Modeling as a method for representing the data on top of which they seek to build their novelty model. In this dissertation, we first describe and explore an educational recommender system called Pique to demonstrate the usefulness of our framework, computational models of novelty, and how they can be applied. Then we show how Topic Models may be the basis for and facilitate measuring novelty of research publications. We describe two different approaches for modeling novelty of research publication data based on Topic Models, one based on atypicality of topic combination, and the other based on similarity, and compare the computational models with human perception of novelty.

## 1.2 Thesis Statement and Research Questions

Applying and considering computational models of novelty as a major part of educational recommender systems can be very useful in encouraging students' curiosity to expand their knowledge. There are not many researches that measure the novelty of research publications based on Topic Modeling. Topic Modeling is developed to automatically generate topics from text; it was not originally developed to measure novelty. In this research, we are suggesting that Topic Models can be the basis for measuring computational novelty in research publications, and that there is more than one way to measure novelty using Topic Models, each of which may look at novelty from a different perspective and result in a different novelty rating. This dissertation states that:

> *Computational models of novelty in research publications augments and complements human perception of novelty in such documents. The computational model of novelty can support learning for students' open-ended projects and for researchers in understanding a large corpus of research publications. Topic Models can be the basis for measuring computational novelty in research publications towards different novelty model approaches. There is more than one way to measure novelty using Topic Models. Atypical combination of topics and similarity are two comparable approaches for measuring novelty of research publications using Topic Modeling. Different novelty models may identify different items as novel showing that*

*novelty has different meanings. Computational models of novelty may differ from human perception of novelty.*

Based on this thesis statement, we can ask the following research questions:

- **RQ1:** How can we characterize the space of possible computational models of novelty in unstructured text documents, and what are alternative approaches to representing unstructured text and computational models of novelty?
- **RQ2:** How computational models of novelty can be useful to encourage curiosity for students' learning in open-ended projects?
- **RQ3:** How computational novelty models of research publications can/may be defined by using Topic Modeling, based on topic combination and similarity approaches?
- **RQ4:** How do topic combination and topic similarity measure novelty differently on the same corpus of research papers? How does the novelty score distribution differ? And, how is the meaning of novelty expressed differently in the two models: co-occurrence (combination) and similarity of topics?
- **RQ5:** How do computational models of novelty compare to human perception of novelty?

Since we do not have any reliable ground truth for modeling the computational novelty of research publications, and there exists a sparse literature on this subject, this research follows an exploratory approach to analyze the results of our investigation and answer the research questions.

## 1.3 Thesis Structure

The structure of this dissertation is as follows: chapter 2 presents our framework for exploring novelty models and existing solutions. It provides a background in modeling novelty and reviews relevant research based on the framework components. Chapter 3 describes the Pique system providing an exploratory study of computational models of novelty in the context of educational recommender systems. In chapter 4 we introduce two approaches for modeling novelty of research publications using Topic Modeling based on 1) combinations of topics (topic co-occurrence), and 2) similarity. We describe the dataset we used and discuss the application and results of the Topic Modeling and novelty models on the dataset followed by an analysis and comparison between the two models. We then compare computational models of novelty with human perception of novelty in this chapter. A discussion about this research and future plans are provided in chapter 5.

# CHAPTER 2: A FRAMEWORK FOR COMPUTATIONAL MODELS OF NOVELTY

Some studies considered novelty as a major component in evaluating creativity regardless of the domain of the data [2–4]. Novelty may also be considered in recommender systems in any domain to recommend novel and interesting items to the user [9, 13, 16–18]. Novelty arises from a comparison in a descriptive space in some cases such as finding the distance of two points in the space [4]. To extend our understanding of computational novelty we establish a framework that facilitates exploring and categorizing novelty models and existing solutions. Our framework shown in Figure 2.1 consists of three major components: 1)type of source data, 2)representation method, and 3)novelty model. Each of these characteristics has a considerable outcome on the performance of intelligent systems. Deconstructing and exploring novelty in this way allows us to have a better understanding of different approaches across domains. In this chapter, we discuss the components of this framework and review relevant research in modeling novelty with a focus on novelty in unstructured text data. We review relevant research from the perspective of the three components of our framework, although not all references can be described as having all three components. Surprise can be considered as a consequence of novelty, being an observer's reaction to novelty, and it has been argued that the same computational models may be applicable to modeling both [17]. Novelty and surprise can be incorporated into recommender systems, with the goal of driving user adoption of new material and thus the broadening of users' preferences [9, 16, 18, 19]. Chapter 2 addresses the first research question concerning how we can characterize the space of possible computational models of novelty in unstructured text documents, and what are alternative approaches to representing unstructured text and computational models of novelty.

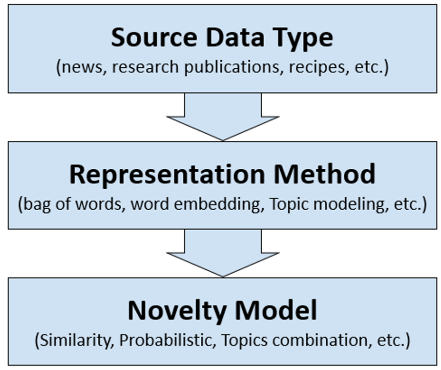



Figure 2.1: A framework for exploring computational models of novelty.

## 2.1 Types of Source Data for Modeling Novelty

The first component of our framework is the type of source data. By source data we mean the raw input data which we want to analyze in order to extract some semantics to use in the novelty modeling and measurement. Considering the type of input data is very important as it can affect the way we measure their novelty. For example, news data are different from scientific data in several ways from the novelty perspective however they are both a kind of textual data: 1)we may have repetitive content in different news documents while scientific papers are generally different from other papers in a corpus. 2)News is not typically in-depth but scientific papers are. 3)Novelty in news is time sensitive while a paper written 30 years ago may still be novel. 4)News topics are not very dependent on prior knowledge but scientific data are. Exploring novelty in textual data can be from different perspectives and in any level of text units (including word level, sentence level, document level, etc.) depending on the source of the data and the approach for the novelty model. In this section, we review some of the most well-known types of textual data applied in various studies for modeling their novelty including news, recipes, and research publications as indicative of how to measure novelty in unstructured text documents. All of these source data types have been used as the basis for research in measuring novelty motivated by an interest in supporting people in finding novel items. For instance, people are always looking for novel news, not repetitive ones. Most people often seek and get excited for novel and surprising recipes. Regarding research publications, a novel contribution is always desired. This led to recommender systems in these domains in recent years seeking to include novelty models as one of the major components in their structure (e.g. [11]).

News articles typically report on events that have occurred recently [20]. Information in a news article is not typically dependent on other news so a user can understand it without background knowledge. Verheij et al. [21] present an evaluation of various novelty methods for ranking web news articles. They investigate different methods for ranking news documents based on a novelty metric using the previously reviewed documents [21]. Allan et al. [22] generate a streaming summary of the news topics which are both useful and novel. Niu et al. [16] modeled serendipity by incorporating surprise and value in the model and implemented it in the domain of health news recommendations.

Another domain of the source text data for novelty modeling is recipes. Novelty in recipes enables people to try new and surprising tastes they have not experienced before. Varshney et al. [23] describe a computational creativity system for culinary design that creates new recipes that both fit to the user's taste as well as being novel and surprising. Morris et al. [24] present a system to generate novel recipes and apply it to inspect computational creativity goals. Grace et al. [11] present *Q-chef* for recipe generation and recommending novel and surprising recipes to stimulate user curiosity and diversity in the users' diet using a database of about 100k recipes from web sources. In another study, Grace et al. [13] developed an approach called "Surprise Walk" to show how co-creative systems could direct the users to understand and value artifacts (in their study ingredient combinations) that are too novel for them initially. They calculated novelty and surprise of ingredient combination pairs from the sGlove word embedding algorithm [13, 14]. They used the "Now You're Cooking" dataset with 80k unique recipes shared on the Internet. The authors used the ingredient set and cuisine tags provided in the recipes for their research. Each ingredient after processing is considered as a single feature in their model (e.g. "pepper" or "apple juice") [13]. Most of the works in the domain of recipes have not considered the quantity or instruction data, and only focus on the ingredients for exploring novelty.

A few studies have investigated the computational novelty of research papers which is the focus in this dissertation. Uzzi et al. [25] analyzed 17.9 million papers from Web of Science (WOS) to find the relationship between combinations of prior work in each paper's reference list, and the novelty and citation count (impact) of each paper. They argue that to have novelty with impact, uncommon knowledge (i.e. atypical combination of journal pairs) is not sufficient and it should be balanced with conventional knowledge (i.e. paper/journal pairings with high frequency) [25]. In another study, Carayol et al. [26] define a measurement of the novelty of scientific articles based on the frequency of pairwise combinations of author-defined keywords and apply it to about ten million research articles published in journals in Web of Science (WoS) during years 1999 to 2013. They also study the relation between team characteristics and novelty, and inspect the forward citations of novel research [26]. In the next section, we review some relevant research from the aspect of data representation method, that is the second component of our framework.

## 2.2 Data Representation Methods for Measuring Novelty

The second piece of the computational novelty framework is the data representation method to be used in novelty models. The source data needs to be represented in some way so that novelty models can be built on top of them. Representation model is important as it provides a basis for measuring novelty. In other words, the representation method provides a bridge between raw data and the novelty model. It provides a processed and structured version of the raw data that can be efficiently used to measure novelty. A commonly used representation method for the source data is representing each item as a vector of features. So each item $I$ is represented as: $I : \{f_1, f_2, \ldots, f_n\}$ where $I$ has $n$ features and $f$ is the value of the item $I$ in each dimension. Representing textual data can be performed in the word level (e.g. word embedding), sentence level, or whole document level (e.g. Topic Modeling). Building meaningful text representations can be challenging. As an example, lexical composition of a phrase may change the meanings of the constituent words and suggest implicit information [27]. In this section we discuss methods for representing text data to be used for novelty calculations, including Bag of Words and TF-IDF Models [28, 29], Word Embedding [14, 30], and Topic Modeling [1, 31].

In the Bag of Words (BOW) model [29], a text document is represented as a collection of words, ignoring grammar and word order. Each text document is represented by a numeric vector with each element being a distinct word of the corpus. The value of each dimension/element of the vector is the frequency of the word in the document, the occurrence by specifying 1 or 0, or it can be another weighted value [32]. In the *Q-chef* study by Grace et al. [11], each recipe is represented as a binary vector of all ingredients in the dataset. Here we can think of each recipe as a text document, and all the available ingredients in the dataset as inspectable words in the dataset. In a study for discovering news articles with the most novel information, Gabrilovich et al. [33] extend the BOW representation with named entities in the text to represent news documents. Some researchers apply a modified version of the BOW model in their studies for modeling novelty of research publications, which uses "bag of keywords" rather than "bag of words" for representing each article [9, 26]. In this approach, each article is initially represented as a bag of author-defined keywords.

TF-IDF [28], short for "Term Frequency Inverse Document Frequency", is another method for text data representation. Using the bag of words model on large corpora may lead to some problems as the feature vectors are based on word frequencies and some words may appear frequently across all documents which can reduce the importance of other words in the feature vector [32]. The TF-IDF model resolves this problem by using the IDF (inverse document frequency) as a normalizing factor. In TF-IDF [28], the term frequency (TF) of each word in the document is multiplied by the inverse document frequency of the word (IDF). The TF-IDF representation method is helpful to discover novel words in the document. Wang et al. [8] applied an approach to measure the novelty of an idea based on TF-IDF by summing the TF-IDF values for all the terms in the idea. Karkali et al. [34] describe a novelty detection method in news document streams based on IDF scoring by applying a TF×IDF weighting model. They keep a summary of the set of observed documents considering the frequency of each term and apply the IDF of each term for a new document and then compute its novelty score using IDF [34]. Compared to the BOW model, feature vectors for the text documents in the TF-IDF model have more scaled and normalized values [32].

Another method applied for the representation of text data is Word Embedding. A word embedding is a learned representation for words and documents based on how the words are used and in which context, such that similar words are close in the vector space [30]. It maps words or phrases to real number vectors with a few hundred dimensions in a continuous space. The theory of the word embedding approach states: "*words that have similar context will have similar meanings*" [29]. The idea of defining the meaning of words by their usage is also stated by Firth [35] "*You shall know a word by the company it keeps!*" [30, 36]. Word2Vec and GloVe (Global Vectors) are two of the main word embedding methods/algorithms for representing text data.

*Word2Vec* [37, 38] is a statistical technique to learn word embedding from a text corpus that is an example of a neural network trained for creating linguistic contexts of words in a new vector space [36, 39]. It was developed and applied in developing pre-trained word embedding by Mikolov et al. [37]. The model constructs a hundred dimensional vector space from a corpus of text and assigns a vector to each distinct word [39]. In this space, the vectors of words with common contexts in the corpus are close to each other [37,38]. This allows vector-oriented reasoning based on the offsets between words. In a *Word2Vec* model a context is defined by a window (a configurable parameter of the model) of neighboring words to learn about words given their usage context [36].

The *GloVe* (Global Vectors) algorithm developed by Pennington et al. [14] is an extension to the *Word2Vec* method for learning word vectors more efficiently. Conventional models for words in a vector representation apply matrix factorization techniques like LSA (Latent Semantic Analysis) which are effective in using global text statistics but are not effective enough as learning methods like *Word2Vec* in considering the meaning and analogies [36]. The *GloVe* technique combines the global statistics of matrix factorization methods like LSA with the context based learning in *Word2Vec*. *GloVe* builds a word co-occurrence matrix by applying statistics in the whole text corpus instead of applying a window to specify the local context as in *word2vec* approach [36]. As stated by Pennington et al. [14], "*GloVe, is a new global log bilinear regression model for the unsupervised learning of word representations that outperforms other models on word analogy, word similarity, and named entity recognition tasks*".

In the *Surprise Walk* study, Grace et al. [13] used a word embedding algorithm called *s-Glove* which is an extension of the *GloVe* model to represent each ingredient as a vector of numbers with 64 dimensions. An advantage of representing each word as such a vector is that similarity (or distance) between words can be easily obtained. They applied this representation for recommending sequences of increasingly novel and surprising recipes. In using the *s-GloVe* vector model, vector subtraction between all pairs of ingredients builds a space (called a surprise space) of combinations of ingredients each with a location and a surprise/novelty rating. In this space "*proximity implies similarity between why those combinations are novel and surpris-*

*ing*" [13]. In [9], the same *s-GloVe* word embedding technique is used to generate a word embedded model using the author-defined keywords of research papers to transform the papers into the same vector space with the goal of recommending interesting papers to students [9].

One of the main disadvantages and limitations of the previous word embedding models is that they cannot properly model polysemous words, that is the words with multiple possible meanings [40]. Existing word embedding models usually ignore polysemy and represent each word with a single vector. Kekec et al. [40] developed a new word embedding model that represents polysemous words by automatically learning multiple representations for each word. A polysemy aware representation gives a more natural embedding of the words and helps to disambiguate word meaning by separating meanings into different maps [40]. Considering polysemy in data representation can lead to a more meaningful novelty model.

Topic Modeling is another approach for representing text data. Among many different representation methods such as bag of words and TF-IDF, word embedding, etc.,Topic Modeling is a less examined approach in modeling novelty of text data, in particular research publications which is our focus in this dissertation. It scans a corpus of text documents and automatically extracts the main topics in the corpus. Topic Modeling provides a much smaller vector dimension compared to the other approaches for representing data such as "bag of words" and TF-IDF vector representation. In those approaches, we have a high dimensional and sparse document representation. If we think of each document as a vector having every word as a dimension and the vocabulary size ($V$), for example, 50,000 words, we will have a vector of about 50,000 of words in "bag of words" or TF-IDF representation. It would not be very efficient to do the novelty measurements if we had about 50,000 dimensional vectors. Topic Modeling reduces the dimension of the vector of variables in large dimensional spaces and gives a meaningful structure for representing text documents. LDA (Latent Dirichlet Allocation) [1, 31] is one of the most basic Topic Modeling algorithms. In [16], Niu et al. used LDA in their second surprise calculation method to discover the themes in the health news articles; however, in their first approach, the topic labels were assigned to the news articles by a human expert, not Topic Modeling. Wang et al. [8] also measure the novelty of crowdsource ideas by applying the LDA Topic Model to represent the vector describing the topic distribution of the ideas. Towne et al. [41] applied the same method but for measuring the similarity (rather than novelty) of about 10k online ideas submitted to the 2012 President's SAVE (Securing Americans' Value and Efficiency) Award. LDA does not take the correlation between topics. In this dissertation, we use STM (Structural Topic Model) [42], as a basis for representing research publications for modeling their novelty. STM models the correlation between topics, which is not considered in the LDA approach applied in previous researches. The assumption in basic Topic Model algorithms like LDA is that all the topics in a corpus are independent and thus no pair of topics is more likely to occur together in a document than the others. STM topic model is not based on this preliminary assumption made by basic Topic Model algorithms. STM considers correlation between topics which is contributory in modeling novelty. We will discuss the Topic Modeling approach in more detail in Section 4.2.

Data representation method is important as the novelty model is built on top of it. Approaches such as Bag of Words and TF-IDF are more helpful to discover novel words in the document, which is more appropriate for novelty detection in news documents. However, for modeling the novelty of research articles, representing data in word level to discover novel words or combinations of words, does not help a lot. Instead, discovering new combinations of topics is more substantial and contributory. To model the novelty of scientific articles, we need a higher level of abstraction like Topic Modeling for representing data (scientific articles). Topics and keywords are two prominent features of scientific papers which can be applied in data representation for their novelty calculation (we will discuss this in more detail in sections 3.2.2 and 4.2). Novelty models are built on top of the data representation model. In the next section, we review relevant research from the aspect of novelty models.

## 2.3 Novelty Models

Modeling novelty is the last and most important piece of our framework which is our ultimate goal too. There have been different definitions for novelty based on the different domains and perspectives. Novelty is typically defined as a measure of the difference between an item and a collection of the other items [2]. A good measure of novelty can help us in intelligent and educational systems to recommending more curiosity-stimulating contents to users and inspiring creativity. Various approaches to model novelty of text data are proposed in the literature including similarity based measurements, probabilistic and information theory based models, frequency, and features combination approaches as some examples [11, 13, 15]. However, few investigated the novelty of research publications as one type of unstructured text data. Also, few have considered Topic Modeling as a method for representing the data on top of which they seek to build their novelty model.

Novel items can be thought of as ones that are not similar to the ones seen/experienced before. Text similarity involves applying a similarity or distance-based metric for finding how similar a text document is with other documents considering features derived from the documents [32]. In the *Surprise Walk* study [13], the use of *s-GloVe* algorithm facilitates measuring the similarity between recipe ingredients pairs. Each ingredient representation in vector space is constructed in such a way that using cosine similarity as the distance metric, the most similar neighbors to it are those ingredients that are surprising in combination with the similar set of ingredients. "Surprise space" is defined as a space of combinations of concepts (ingredients) with a location and a surprise score for each combination, and "*proximity implies similarity between why* *those combinations are surprising*" [13]. A list of the combinations, which increase in surprise score monotonically from user familiarity towards the novel target surprise, is generated by using cosine similarity as the distance metric between combinations vectors. In Gabrilovich et al. study [33] of discovering news articles with most novel information, the authors applied various metrics including *Kullback-Leibler (KL)* divergence and *Jensen-Shannon (JS)* divergence [43] to measure difference between each news article and the set of previously read ones for identifying the most different and novel articles. They used a sliding window for including a number (in their study 40) of prior articles to assess the novelty of each new article. They discuss that their methods for document similarity to identify novel articles can fulfill the requirement of personalized news portals and news alerting services for reducing the time and disruptions to the users of evolving news stories [33].

Frequency is another perspective for measuring novelty. Typically items/events that occur frequently seem less novel to the observer. That is also the case for the frequency of word occurrences in text. Word frequencies are typically associated with various text properties such as novelty of text [44]. Carayol et al. [26] apply the frequency of pairwise combinations of author-defined keywords in measuring the novelty of scientific articles. Karkali et al. [34] proposed a method for novelty detection in news document streams by applying IDF scoring. They argue that applying IDF in novelty modeling provides a faster novelty calculation compared to other similar studies because there is no need to compare the similarity of a new document to all the prior documents in the text stream. They describe that a novel document typically uses different words compared to the words in the prior documents, and conclude that the words of a novel document typically have high IDF scores as they have high specificity [34]. One disadvantage of their method as they discuss is that their novelty model cannot capture synonymousness and that can be an issue in applying bag of word representation and vector space model. That is, a document with words synonymous to the words in the other documents may be recognized as novel for having words with high IDF values while it is not novel [34].

Another perspective in modeling novelty is considering combination of features in the items. Atypical and rare combination of features in an item indicates the novelty of that item compared to the rest items. This type of novelty measurement is also borrowing the concept from similarity and frequency approaches. By considering the frequency of co-occurrence of any two (or more) features, novelty can be defined as any rare (new) combination of features that is not similar to the past observed frequent combinations. Morris et al. [24] measure novelty in a recipe by counting new combinations of (as they described "known") ingredients or rare n-grams. An n-gram is a combination of n ingredients. For example, a 2-gram could be wine, garlic. They explain in their study that a rare n-gram is an n-gram that does not occur frequently and does not contain a rare (n-1)-gram as a sub-combination (for instance, 4-grams should not contain rare 3-grams) [24]. Uzzi et al. [25] analyzed 17.9 million papers from Web of Science (WOS) to find the relationship between combinations of prior work in each paper's reference list, and the novelty and citation count (impact) of each paper. They argue that to have novelty with impact, uncommon knowledge (i.e. atypical combination of journal pairs) is not sufficient and it should be balanced with conventional knowledge (i.e. paper/journal pairings with high frequency). In Niu et al. study [16], the first surprise calculation method (which is a variation of Mutual Information) considers each Health News article as "a bag of co-occurring topics". However, the topics in this method are not derived by applying Topic Modeling but are the labels assigned to an article by experts. They discuss that a rare topic combination in an article gives a higher surprise score for that article. In Carayol et al. [26], authors propose a measurement of the novelty of scientific articles based on keyword pairwise combination frequencies that is computed on the set of research articles that have at least two keywords in the WoS over fifteen years (from 1999 to

2013). In this dissertation, we consider combination of topics extracted by applying Topic Modeling to model novelty of research articles. In contrast to the studies in which a human expert assigns topic labels to the document (e.g. first approach in [16]), or the author assigns keywords to a single paper (e.g. [26]), Topic Modeling automatically provides consistent topics by scanning and considering all the corpus consistently, not just by some human expert's opinion. In Topic Modeling, there is consistency in the identification of features across the entire dataset, whereas author-defined keywords provide features relevant to the author of a single item/article in the corpus which lacks consistency.

Some studies consider information theory and entropy as the metric for measuring novelty by computing the information content of a dataset [15]. Entropy in information theory is a measure of the uncertainty that is associated with a random variable. An entropy function can be applied by researchers to gauge the level of disorder of the remaining dataset after removal of points with high entropy which are considered as novel [45]. It is assumed that novel data contain more information to convey and consequently make the observer surprised [15]. As stated by the Shannon [46] definition, the amount of information contained in a piece of data $D$ is measured by $-\log_2 P(D)$ bits. That is a rare piece of data with small probability has more information to convey [15]. One way to measure the amount of information and novelty in a piece of data is to see how much the observer gets surprised by observing that piece of data, that is "*measuring the difference between the observer's prior and posterior belief distributions*" (i.e. the effect of the data on the observer) [15]. Baldi and Itti [15] use relative entropy or Kullback-Liebler divergence [47] as one way of measuring the surprise (distance between posterior and prior distributions [15]). By measuring the surprise we can also conclude about the level of novelty of the data. In their study [15], an observer is defined in terms of Bayesian statistics, who has a probability distribution (prior) over what he thinks will happen that is updated by using Bayes theorem whenever they receive new information. The updated distribution is called the posterior. The authors define a unit of surprise, the "*wow*", as $-\log_2 P(M)/P(M|D)$ where "$P(M)$ *is the observer's prior distribution and* $P(M|D)$ *is the observer's posterior distribution after observing data* $D$" (or an event) [15]. They discuss that this "*bayesian surprise measures a facet of information that is different and complementary to Shannon's definition*" [15]. This approach is very different to some of the others because what's being measured is the observer's beliefs, not anything in the actual world.

In Varshney et al. [23] work, the novelty of a new recipe is calculated using Bayesian surprise defined as the *Kullback-Leibler* divergence method quite similar to the method presented in Baldi and Itti [15]. In the *Q-chef* system [11], Grace et al. implemented the model of surprise using the *wows* method described in [15] and calculated the likelihood of an ingredient given a set of other ingredients. They explain that one *wow* of surprise shows that one extra bit of information is given by occurring that feature in that context, that is it is half as likely to occur [11]. In the *Surprise Walk* study [13], a surprise score for each pair of ingredients is also calculated with the same method in [11] inspired by Baldi and Itti [15]. The authors describe the *surprise space* to generate interesting suggestions for directing users toward more novel content (ingredient combinations/recipes). The target artifact in their work is a novel combination of ingredients with a high surprise score. Given that, a list of the combinations (which increase in surprise score monotonically from what the user is familiar with, to the novel target surprise) is generated by using cosine similarity in the process as the distance metric between combinations vectors (see [13] for a complete explanation). By recommending this list of ingredient combinations the user is avoided to be overwhelmed by facing the target novel combination and can appreciate it at the end of the suggested sequence [13]. In [23], the novelty of a new recipe is calculated using Bayesian surprise defined as the *Kullback-Leibler* divergence method quite similar to the method proposed in [15].

In this chapter, we provided the answer to the first research question as, *RQ1: How can we characterize the space of possible computational models of novelty in unstructured text documents, and what are alternative approaches to representing unstructured text and computational models of novelty?* We defined a framework to explore and characterize different novelty models in unstructured text. We also reviewed some of the remarkable alternative approaches from the perspective of each component of our framework. A novelty model that is appropriate for a specific domain of textual data and representation model, is not necessarily suitable for another one. For example, detection of novel documents in text (news) streams using similarity (distance) metrics can be slow despite its accuracy, as an incoming document should be compared with all the previously observed ones. In recommender systems, typically items are recommended to the user that are in very close match to the user's preferences and knowledge, which can cause fixation. To overcome the challenge of fixation in educational recommender systems, infusing novelty to find and recommend novel items and learning materials to the students is very important. In the next chapter, we demonstrate the usefulness and role of computational models of novelty and our computational novelty framework in educational recommender systems, by describing and exploring Pique as a web-based educational recommender system. We report on applying computational models of novelty in educational recommender systems to encouraging students' curiosity. Pique as a primary study represents how our computational novelty framework can be the basis for the AI in educational recommender systems like Pique.

# CHAPTER 3: AN EXPLORATORY STUDY OF COMPUTATIONAL MODELS OF NOVELTY IN THE CONTEXT OF EDUCATIONAL RECOMMENDER SYSTEMS: PIQUE

Identifying novel and valuable content, designs, and articles can lead to surprising recommendations and consequently stimulate a curiosity to explore beyond what we already know [9, 48]. Theories of intrinsic motivation consider novelty and surprise as two of the main factors that evoke interest, motivate exploratory behaviors, and consequently drive learning and creativity [49]. Chapter 3 addresses the second research question concerning how computational models of novelty can be useful to encourage curiosity for students' learning in open-ended projects?

In this chapter, we describe the development, application, and study of computational models of novelty in educational recommender systems for encouraging students' curiosity. We developed a web-based personalized recommender system called Pique that applies AI-based computational models to identify novel documents from a data set of learning resources, and then generates a sequence of learning materials personalized to an individual's knowledge and interests. This approach enables instructors to set a class-wide task with a fixed corpus of learning materials, but for each student's experiences to be personalized in open-ended student-led and/or project-based learning [50, 51]. To demonstrate how computational models of novelty can be leveraged to encourage curiosity, we describe the Pique system from the perspective of our defined framework for computational models of novelty (see Chapter 2). In Pique, we extended this framework by adding a fourth component of personalization to it for personalizing the learning materials to the students as shown in the AI element of Pique in Figure 3.1 (see the orange framework in Figure 3.1). The updated framework provides an ontology for computational novelty with respect to its four components: the source of text data, methods for representing the data, models for measuring novelty, and personalization [17]. We discuss Pique from the viewpoint of the computational novelty framework and report on students' experience of using Pique.

In Pique project, we present two data representation methods, two computational models of novelty, and three ways to select learning resources to stimulate students' curiosity. The novelty models are based on the concept of unexpectedness as a cause of novelty and surprise, which consequently leads to curiosity [13, 48]. For example, learning materials containing interesting and unexpected information can create a surprise response, which may drive students to explore those concepts further. Presenting students with novel learning resources related to but distinct from their knowledge can inspire their curiosity to explore more in the domain.

The fourth piece added to the framework as its last component for being applied in Pique, is personalization and sequence generation to prepare a set of recommendations for the user/learner in recommender and learning systems. Using the novelty scores assigned to each document (learning resource) by the novelty model from the third component of the framework, an algorithm will be defined to generate a sequence of item recommendations based on the user topic/keyword selection. In Pique, we explored three personalization and sequence generator approaches during the project. All of these approaches are based on the novelty score obtained from the third component of the framework as well as students' keyword or topic selection as their interests. Each of these three models has its own strategy to produce a sequence of learning materials to recommend to the student. These models and related algorithms are described in more detail later in this chapter.

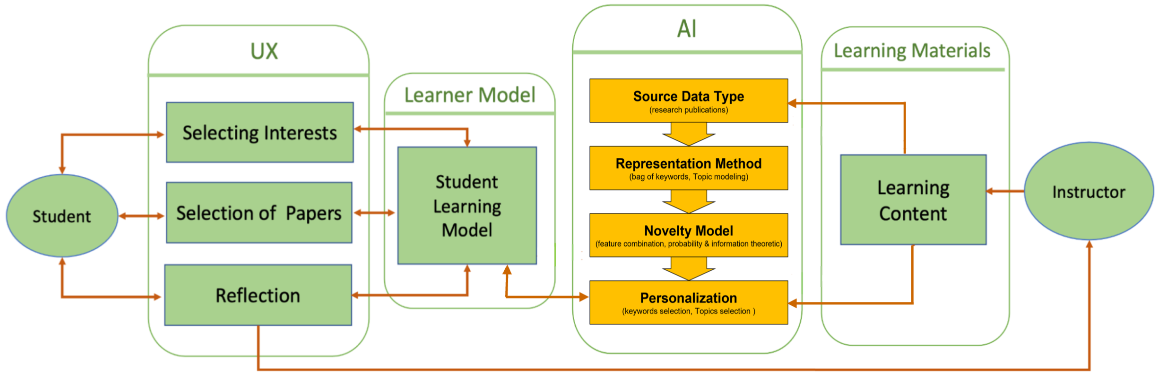



Figure 3.1: Architecture of the Pique Learning System, applying the framework for computational models of novelty indicating the approaches in Pique.

In this chapter, we shed light on the role of computational models of novelty in personalized educational systems such as Pique, and how computational novelty models could be leveraged to stimulate student curiosity and expand their learning interests. We apply our generalized framework for computational models of novelty as the basis for the AI component of the Pique system. We describe the Pique model for encouraging curiosity in learners in a project-based open-ended course experience. The framework is described to provide structure for the use of computational novelty in Pique and is generalized to inspire this approach in other domains and courses. Pique is presented as a model and an implementation, with an evaluation of this approach based on students' experiences with Pique in 2 semesters of 2 different courses. In this chapter, we describe the architecture of the Pique system and its implementation in personalizing learning materials. We identify specific computational approaches to each component used in Pique based on the framework for computational models of novelty, describe the Pique model as well as the development and implementation of the Pique system, and finally report on the experiences of university students who used Pique in the classroom.

Pique as an educational learning system consists of four main elements of learning materials, artificial intelligence methods (AI), learner model, and user experience (UX). As illustrated in Figure 3.1, all of these elements have close interrelation with the components of our computational novelty framework in the AI element. Following we describe different parts of the Pique system as depicted in Figure 3.1, and explore the four components of the computational novelty framework pertaining Pique.

## 3.1 Learning Materials

The instructor provides the source of documents as the learning material for a specific course. The learning material for our deployment of Pique is selected based on its relevance to the courses in which we used the Pique system. We included Pique in two courses in a Computer Science program. The first course titled, "Human Centered Design", has a focus on human-computer interaction. The learning materials for this course are articles published in the ACM Digital Library under the classification of Human-Centered Computing. The second course, titled "Graduate Teaching Seminar", has a focus on educational research in computer science, and the relevant learning materials are articles published in the ACM SIGCSE (Special Interest Group on Computer Science Education) proceedings. For the "Human Centered Design" Course we collected a total of 9,452 conference, journal and magazine papers with publication dates between 2008-2018. For each publication, we extracted the title, ISSN, location, abstract, publisher, address, ACM ID, journal, URL, volume, issue date, DOI, number, month, year, pages, and tags/keywords as metadata. For the Graduate Teaching Seminar, we collected a total of 1172 papers with publication dates between 2008-2018, with the following metadata: title, author, conference, year, DOI, keywords, and abstract.

## 3.2 AI in Pique

The AI element in Pique is the most important part, i.e. the heart of the Pique system which includes the four components of the computational novelty framework described in section 3. In this section, we elaborate on the framework components and the approaches/methods we used in each component for Pique.

### 3.2.1 Source Data in Pique

The source data or the first component of the computational novelty framework in Pique is research publications. As described in section 4.1, the documents in the datasets for the two courses in which we included Pique are unstructured text extracted from conferences, journals, and digital libraries relevant to each course. Following we describe the methods we used in Pique for preparing and representing data to be applied in novelty model.

### 3.2.2 Data Representation Methods for Pique

The approaches to computational novelty are dependent on the representation of the items for which we are measuring novelty. The representation of unstructured text documents plays an important role in achieving an effective novelty measurement. Two representation methods we applied in Pique to represent the source data include Topic Modeling and bag of keywords. Considering the paper keywords and (main) topics as the most prominent types of features for a scientific article, each learning item was represented by extracting a list of features based on keywords or Topic Models associated with each item/document. These features provide the basis for computing a novelty score for each item discussed in the next part.

**Applying Bag of Keywords for Representing Data in Pique.** For the first representation approach, each item of the learning materials is represented as a bag of keywords. With the keywords for each paper, we created a bag of keywords representation for measuring novelty in the next step. Identifying the keywords for the learning materials for each item was challenging in this approach. In the dataset, each paper includes two fields in the metadata that can be considered as the keywords for this model. One is the keywords selected from the ACM's Computing Classification System (CCS), and the other is author-defined keywords. The ACM Computing Classification System is developed as a poly-hierarchical ontology resulting in common topics relevant to all papers, but they do not specifically represent the content in each paper. On the other hand, author-defined keywords are defined for each specific paper without following any standard representation. To make the data representation prepared for the novelty model, we synthesized the list of keywords from each paper into a master list of keywords for the dataset. We then created a mapping from a user's interests to the concepts in the learning materials by manually curating a reduced set that can be used for mapping. Considering too many keywords would be overwhelming, and inadequate keywords would not represent the dataset with enough fidelity, we tried to choose the number of keywords that are reasonable to present to students for selection. We manually replaced keywords that were not in the reduced list to be the most relevant keyword in the reduced set. Across the semesters, feedback from students indicated that our reduced set of 35-55 keywords was sufficient for students to express their interests.

**Applying Topic Modeling for Representing Data.** For the second representation approach, we adopted a Topic Modeling approach for deriving concepts from the corpus. A Topic Model [1, 31, 52], is a type of statistical model for learning and extracting the hidden semantic structures (main topics or themes) that occur in a corpus of text documents. Each extracted topic consists of a probability distribution over all the words in the corpus and each document consists of a probability distribution over the topics [1, 31, 48, 52]. We describe the Topic Modeling representation method in more detail in section 4.2. By applying Topic Modeling algorithm to the corpus of learning materials in Pique, each item of the learning materials is then represented as a vector of topic distributions, and a 20x20 dimensional topics correlation matrix is provided including the correlation coefficient for all topic pairs (see section 4.2 for more details on Topic Modeling technique for representing text data).

### 3.2.3 Computational Models of Novelty for Pique

We implemented two computational models of novelty, based on probability and information theory, and features combination. One novelty model is referred to as the "Keyword co-occurrence model" and the other as the "Topic co-occurrence model". Each item in the learning materials is represented as a bag of keywords in the keyword co-occurrence model, while the topic co-occurrence model applies Topic Modeling approach/method to represent each item in the learning materials as a vector of topic distributions. Following we describe how these computational approaches assign novelty scores to the documents in the corpus of learning materials specified by the instructor of the course.

**Novelty Model Based on Keyword Co-occurrence.** The first novelty model in Pique is based on the probability of co-occurring for each pair of keywords in the corpus. This model benefits from a variation of the *Mutual Information* for calculating novelty as in [16, 53]. Having a bag-of-keywords representation for each paper, we calculated the co-occurrence of keywords for measuring novelty. We removed papers with fewer than two keywords, and then measured the probability of each pair of keywords appearing together in the corpus. We applied the resulting probabilities shown in equations 3.1 and 3.2 below to get the probability of co-occurring of keywords $x1$ and $x2$ in the corpus as shown in equation 3.1. By taking its logarithm we got the novelty score for that pair of keywords. A novelty matrix $NM$ (equation 3.4) was then created for all pairs of keywords in the corpus, considered as the look-up table for identifying the novelty scores among the keyword pairs in the papers. The highest value of all keyword pairs present in a paper was then used to get the score for the paper as surprising combinations stand out [48] which is shown in equation 3.5.

$$prob(x_1) = \frac{\#\ of\ papers\ have\ x_1}{\#\ of\ total\ papers} \tag{3.1}$$

$$prob(x_2) = \frac{\#\ of\ papers\ have\ x_2}{\#\ of\ total\ papers} \tag{3.2}$$

$$prob(x_1, x_2) = \frac{\#\ of\ papers\ have\ both\ x_1\ and\ x_2}{\#\ of\ total\ papers} \tag{3.3}$$

$$NM(x_1, x_2) = \frac{log_2(prob(x_1, x_2))}{prob(x_1) * prob(x_2)} \tag{3.4}$$

$$NoveltyScore\_P_n = max(NM(x_1, x_2), NM(x_1, x_3), ...) \tag{3.5}$$

**Novelty Model Based on Topic Co-occurrence.** The second novelty model in Pique applies the Topic Modeling approach for representing each paper as a vector of topic proportions. This model considers the overall novelty of a document to be equal to the most novel concept or combination of concepts within that material [48]. The novelty of a text document is calculated based on the lowest (i.e. highest negative) correlation coefficient among all pairs of topics significantly present in that document, and the proportion of the document which contains that pair. To determine whether a topic is "significantly present" in a document, a topic proportion threshold of 0.1 is used, that is the document should be at least 10% comprised of that topic. This model is based on our previous work in topic-model approaches to novelty [48]. Equation 3.6 shows the novelty formula for a paper $p$ considering the set of topics significantly present in $p$. The pair of topics in $p$ with the lowest correlation coefficient are denoted by $t_i$ and $t_j$ which are considered as the most novel topic combination in $p$. This coefficient is divided by the correlation of the most novel pair of topics in the whole corpus that are $t_a$ and $t_f$ and then is weighted by the proportions of ti and $t_j$ in p for computing the novelty score.

$$NoveltyScore\_P_n = \frac{CovMat(t_i, t_j)}{CovMat(t_a, t_f)} \times 2(min(prop(d, t_i), prop(d, t_j))) \tag{3.6}$$

In this formula $CovMat$ is the covariance matrix obtained from the topic model, $CovMat(t_i, t_j)$ is the correlation of the document's most atypical topic combination $(t_i, t_j)$, and $CovMat(t_a, t_f)$ is the correlation of the most atypical topic combination in the whole corpus. $prop(d, t)$ is the proportion of document $d$ that consists of topic $t$. The expression in the parentheses is the novelty of the document's most novel topic combination, represented as a proportion of the most novel topic combination in the model. In the next section, we describe the process and algorithms developed in Pique for finding appropriate papers to recommend to students which are both novel and match their interests.

### 3.2.4 Personalization and Sequence Generation Algorithms in Pique

Pique personalizes its recommendations by including student's selection of keywords/topics of their interest in the process of generating the sequence of papers for recommendation. In generating the recommendation sequence, Pique takes the novelty ratings of each document in the corpus and constructs a sequence of learning resources that maximize the chance of a student experiencing optimal novelty. The goal is generating a sequence of learning resources to support student-directed learning and to stimulate students' curiosity about learning. We explored three sequence generator approaches during the course of our project. We named these three models as "Origin-Destination model", "Destination model", and "User-Directed model".

Pique generates a personalized sequence of nine documents in sets of three papers from the corpus of learning resources based on student information and preferences. Students select one paper from each set of three, read it, and reflect on it. Then students are presented with the next set. The different sequence generator models are based on different representations of student interests. The Destination Model uses one set of student-specified interests as the input to the algorithm. In the Origin-Destination Model two student-specified sets of keywords are used: one that they self-report as already knowing about or the "origin" set, and one that they want to learn more about, or the "destination" set. The User-Directed Model extends the Origin-Destination Model to include other keywords from the papers most recently selected by the student. The sequence generator uses these keywords to represent student preference and combines that with the novelty score for each paper to select and sequence learning resources with the goal of inspiring curiosity.

**Destination Model.** The Destination Model asks for what students desire to learn and recommends a set of nine novel documents relevant to their stated desires. When applying with our keyword co-occurrence novelty model, the student interests are directly mapped to the corpus keywords, but in the case of applying the topic co-occurrence model, a mapping was manually built between the topics automatically generated by the topic model and the keyword set we had created. Here we refer to "novel documents" generally, without specifying which novelty model labeled them as such. Initially, students select their learning interests, which is considered as the destination set, $D$. Then the destination model identifies documents in the learning materials corpus for which the top $N$ topics within that document include at least one of the user's selections. We decided on $N = 3$, as we found most documents in the corpus included at least this many topics at reasonable proportions. From the set of identified documents, the nine most novel papers are selected and sorted ascending based on their novelty score, from the moderate novel to the most novel one. The goal in the Destination Model is recommending nine documents with information that students want to learn, starting with a moderate novel document and then scaling up to highly novel documents as the student reads more and learns about their interested topics.

**Origin-Destination Model.** The goal of the Origin-Destination model is inspiring students to explore learning materials that contain some information that they already know, combined with some new information that they don't. New material is better learnable if it is somewhat connected to topics already known [54]. This model generates a recommendation sequence that moves from what the student already knows to what they want to know. The algorithm is inspired by the *surprise walks* algorithm [13] that moves from an unsurprising source to a surprising destination in the recommendation sequence.

In the Origin-Destination Model, the learning materials are presented in three steps of "close", "far", and "farther" to stimulate learners' curiosity. Recommending the learning materials step by step helps students to gradually learn new materials similar to what they already know and inspire them to explore without recommending materials that are so novel as to be unfamiliar and overwhelming for them [55]. In the first step, papers similar to student's familiarity, which are labeled as the "close" category of learning materials for that student, are recommended by the model. In the second step, papers that are similar to both what the students already know (their familiarity) and what they want to learn, labeled as the "far" category, are recommended by the model. In the third step, papers containing materials related only to what students want to learn, labeled as the "farther" category, are recommended to student by the model.

For the "close" category, the model identifies candidate papers containing at least one common keyword (or topic) from the students' initial interest set (source set). By applying the k-means algorithm the model clusters the candidate papers based on their novelty scores to distinguish the papers with three novelty levels of high, medium, and low. The model computes the paper's familiarity score as well, denoting the number of keywords in common between the paper and the "origin" set of keywords/topics the student already knows. Then in each novelty level, the papers with the highest familiarity scores are selected, and finally, the algorithm recommends one low, one medium, and one high novelty paper. Regarding the "far" category, the model recommends another three papers for expanding students' learning from what they are familiar with to the new topics they would like to learn. Candidate papers in this category include at least one common keyword from the origin keywords set ($O$) and at least one common keyword from the destination keywords set ($D$). The same clustering approach is applied to identify low, medium, and high novelty candidate papers, and the candidate papers are identified in each level with the highest number of common keywords. For the "farther" category, papers containing information that students desire to learn are presented by the model. Candidate papers in this category include at least one keyword from the destination keywords set ($D$), and similar to the other two sets the candidate papers are categorized into three levels of novelty.

**User-Directed Model.** User-Directed model extends the Origin-Destination Model by considering students' decisions during the recommendation process in order to recommend materials aligned with their evolving interests. As in the Origin-Destination model, the papers are recommended step by step by the three categories of close, far, and farther, but this model additionally keeps track of students' selections of papers from the previous step. Keywords of papers in the previous step are applied in order to prioritize similar resources in the recommendations of the next step. The User-Directed model filters the candidate papers for the far step to those that share at least one keyword with papers selected in the close step. The model first identifies candidate papers for the farther step that contain at least one keyword in common with the keywords of the paper selected in the far step. This model is identical to the Origin-Destination model except for the aforementioned filtering step. That is, it recommends one low, one medium, and one high novelty paper in each of the close, far, and farther steps.

## 3.3 Learner Model

The Learner Model in Pique collects information about the learner to support the selection and presentation of learning materials plus information needed to analyze the use of Pique. It is in the direct connection to the personalization component of the framework in the AI part. The Learner model is not a comprehensive model of the learner. It stores two kinds of information: information about the students and how they have used Pique to date. Most of the information about students remains constant as their name, ID, email address, and course. The IDs are automatically generated by the Pique system and serve to de-identify students as required by our IRB approval. The information in the student's profile that can be changed is their interests, which they select when they start using Pique but are prompted to change for each recommendation cycle. Student's data is updated with a new cycle record every time the student uses Pique. The cycle records include timestamps, the papers they selected, the options they chose from, and their reflections. Their reflections include their responses to three questions: 1) why the student has selected the paper, 2) if the selected paper is on a topic matching their interests, and 3) what topics the student expects to learn from the paper. These reflections are collected after students read the papers and are used for the research.

## 3.4 UX for Pique

The User Experience of Pique supports students' interaction with the following three steps: Selecting interests, Selection of papers, and Reflection. We designed the student experience of using Pique as a cycle of recommendation followed by reflection. The students log in so that Pique can track their selection of topics/keywords, their selection of papers, and their reflection on the papers they read over the course of the semester.

**Selecting Interests.** Pique captures students' interest by prompting them to identify what they want to know. This prompt assists students to formulate their learning goals and provide them with more control over their learning choices and enables self-directed learning. After students log in to the system, they are prompted to select the topics/keywords they would like to know about from a checkbox interface. Figure 3.2 shows the user interface with the learning options for the students as they were in the Graduate Teaching Seminar course.

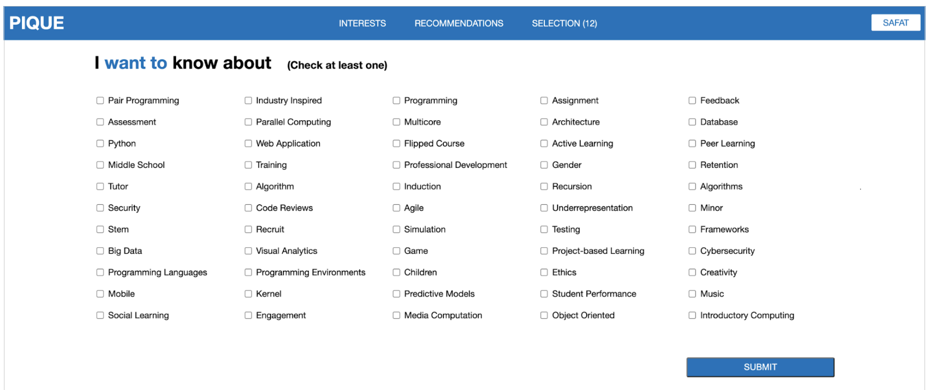



Figure 3.2: UX for selecting interests in Pique [18].

**Recommendation and Paper Selection.** After students submit their selected topics/keywords, they are navigated to the recommendation page (Figure 3.3) showing a sequence of nine papers in the area of their selected topics/keywords. For each paper, the students are presented with the title and novelty score of the paper, and papers are sorted by their novelty score. Students can view and download the pdf file of the paper by selecting/clicking it. This step of paper selection in Pique enables students' self-regulated learning, with the intention of stimulating their intrinsic motivation to learn and explore. This stage presents the papers that are recommended by the student selection and sequence generation of Pique (the fourth component of the framework in section 2). Pique presents the nine papers in sets of three, based on the sequence generation algorithm (see Section 3.4). Figure 3.3 shows an example of papers being recommended in the Graduate Teaching Seminar course based on the Origin-Destination sequence composition model. The top three papers are closely related to what the student already knows, the middle three are related to both what they know and what they are interested in, and the bottom three are related to what they are interested in only. The step of paper selection informs students about how novel a particular paper is and allows them to manually choose more or less novel papers by selecting the drop-down menu labeled "show me papers" in the top right corner of Figure 3.3. In this way, students have the option to explore a wider range of papers in their selected topic/interest category.

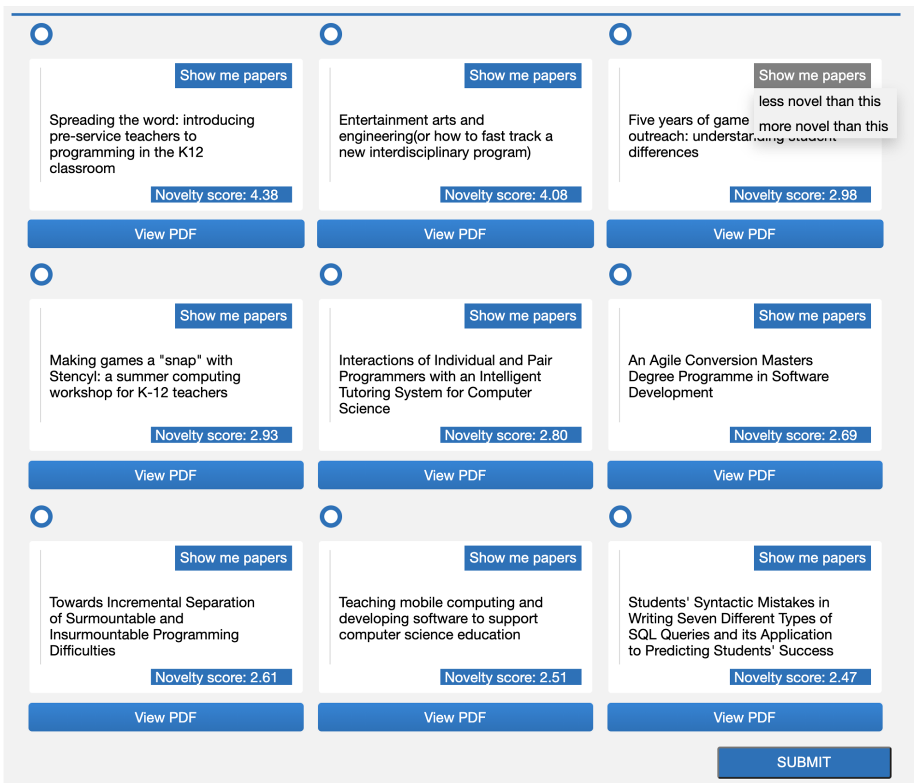



Figure 3.3: UX for recommendation sequence, and selecting learning content based on interests and novelty scores [18].

**Reflection.** The third step of the Pique UX is Reflection. It has been shown in cognitive studies of students that reflection is key to effective learning [56–58]. There are two types of reflection in the Pique system. One is requested when students select a paper to read as shown in Figure 3.4, and one is requested at the end of the semester. The first reflection asks the student to answer 3 questions about the paper they selected (Figure 3.4). The first question asks about why they selected this paper. The second question asks whether the selected paper is on a topic the student expressed interest in, and the third question asks about what the student expects to learn from this paper. After completing the survey, the student can log out or continue to the next round of the recommendation cycle. The second type of reflection asks students to reflect on their overall learning experience. Students were asked to summarize the papers they read and categorize those papers into groups. Students are asked to identify the paper they found most interesting and justify why. This reflection allows students to organize their newly acquired knowledge where the learning paths are constructed by the students rather than the instructors. It was also critical for evaluating the impact of this educational innovation on the student experience.

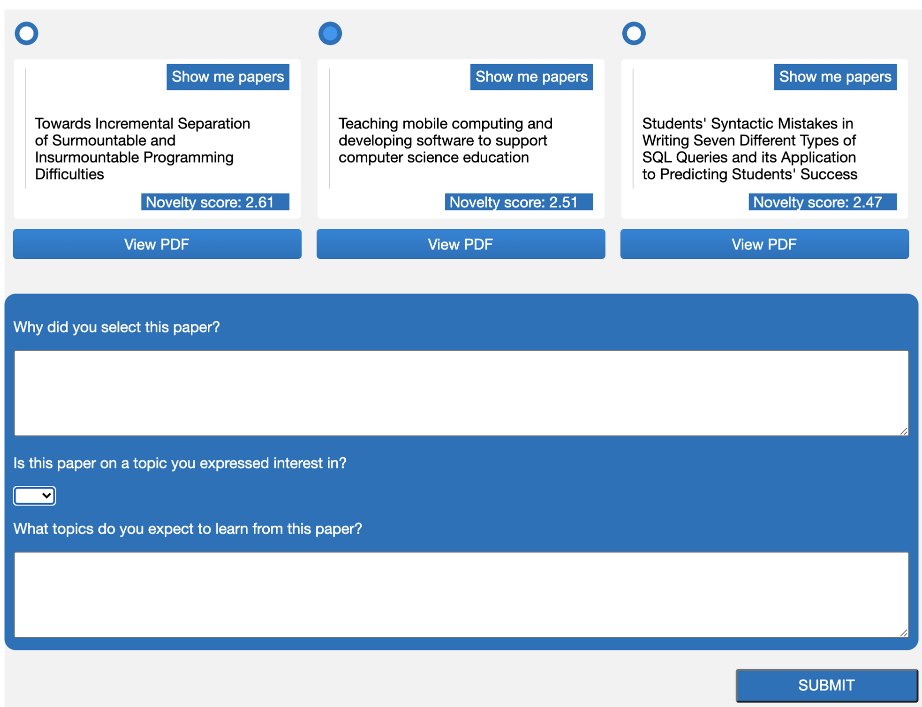



Figure 3.4: Pique asks students to reflect on their paper selection and learning Expectations [18].

## 3.5 The Student Experience Using Pique in Specific Courses

We applied Pique in an undergraduate human-centered design (HCD) capstone course and a graduate teaching seminar course for PhD students. These 2 courses are project-based, where the HCD course requires research relevant to a design project and the graduate teaching seminar has focus on reviewing research for a project report on graduate teaching. We used Pique over several semesters and continually developed the models of novelty and sequence generation based on student and instructor feedback. The goals of this study are to answer the following questions based on students' experience with Pique and their reflections on the recommended learning content:

- Goal 1: How does the experience of using Pique enable self-directed exploration and personalized learning?

- Goal 2: How does the experience of using Pique assist students in expanding their learning interests?

In this section, we describe the deployment of Pique and the experiences of students who have used Pique in the classroom through a quantitative and qualitative analysis of student data collected during the course experience. We have IRB approval for the data collected by Pique.

### 3.5.1 Quantitative Analysis of Students Experience

We used Pique over four semesters in both undergraduate and graduate courses in Human Centered Design as well as the Graduate Teaching Seminar PhD course. In the Human Centered Design course students were asked to use Pique for six weeks, and had to submit weekly and end-of-semester writing assignments about the papers they had read. Each week they were asked to submit a summary of the three papers they downloaded and read, and identify the most interesting paper among the three. For the end-of-semester report, the students were asked to explain their experience of using Pique, what they learned, the most interesting paper they found among all, and their reason for why they found the paper the most interesting. For the Graduate Teaching Seminar course, students were asked to use the Pique system for the entire semester but submitted only a final report without any weekly submissions. This was due to the PhD students' greater familiarity with reading published articles, as well as their overall greater autonomy as learners.

Regarding our first research goal concerning how the use of Pique assisted in enabling self-directed exploration, we investigated how the student cohort differed in the resources they explored, as a measure of how self-directed their experiences were. Table 3.1 shows the summarization of our results. Though students' options for selecting interests stay the same for all cycles, that is 39 interests in Human Centered Design and 55 in Teaching Seminar course, we found that students were presented with very diverse sequences of learning resources. A total of 621 unique papers were recommended by Pique in the Graduate Teaching Seminar course for one semester, even though this course included just five students. The results showed 55% of those papers were recommended to at least two students, due to overlaps in topics of interest. Those five students selected a total of 66 papers to read, showing 86% of the selected papers were selected by just one student. Across all four courses, we observed 72% of recommended papers were recommended to at least another student, but the students' selections were highly diverse, showing 70% of the selected papers were unique to that individual student.

Our second research goal asked how using Pique helped students in expanding their learning interests. We investigated the change in students' interests overtime for responding to this question as illustrated in Figure 3.5. The top two charts in Figure 3.5 are related to the HCD courses in Spring and Fall semesters. The bottom two charts are related to the Graduate Teaching Seminar courses in the Spring and Fall semesters. In all semesters in which Pique was used, we observed

Table 3.1: Distribution of learning materials to personalize learning

| Course name | Graduate Teaching Seminar | Graduate Teaching Seminar | Human Centered Design | Human Centered Design |
|---|---|---|---|---|
| Semester | Spring 2020 | Fall 2020 | Spring 2020 | Fall 2020 |
| Number of students | 24 | 5 | 12 | 12 |
| Number of unique learning sequences generated by students | 24 | 5 | 12 | 12 |
| Total papers selected by students over the Pique cycles | 221 | 66 | 76 | 77 |
| % of selected papers uniquely picked by individuals | 50% (111 papers) | 86% (57 papers) | 71% (54 papers) | 74% (57 papers) |
| Total papers recommended by Pique | 1987 | 612 | 729 | 774 |
| % of papers recommended to at least one other | 84% (1669 papers) | 54% (333 papers) | 75% (548 papers) | 72% (558 papers) |
| % of papers recommended to only one student | 16% (318 papers) | 46% (279 papers) | 25% (181 papers) | 28% (216 papers) |

an increase in the interests selected by the students. The X-axis shows the number of Pique cycles and the Y-axis shows the average cumulative growth of the interest selections. We computed the number of interests selected by each student for each cycle. We aggregated this for all students within a cohort to give the average number of interests selected by the students in that cycle. The cumulative number of interests in Figure 3.5 demonstrates the expansion of stated interests over the semester. The total students in the HCD courses selected an average of only four interests at the beginning of applying Pique. As students used the Pique system over the semester, we observed that searching of learning interests increased as well. At the end of the semester, all students had explored an average of 67 interests. Regarding the Graduate Teaching Seminar, total students started with just two interests on average, and over the semester the average number of their searching learning interests raised to 42.

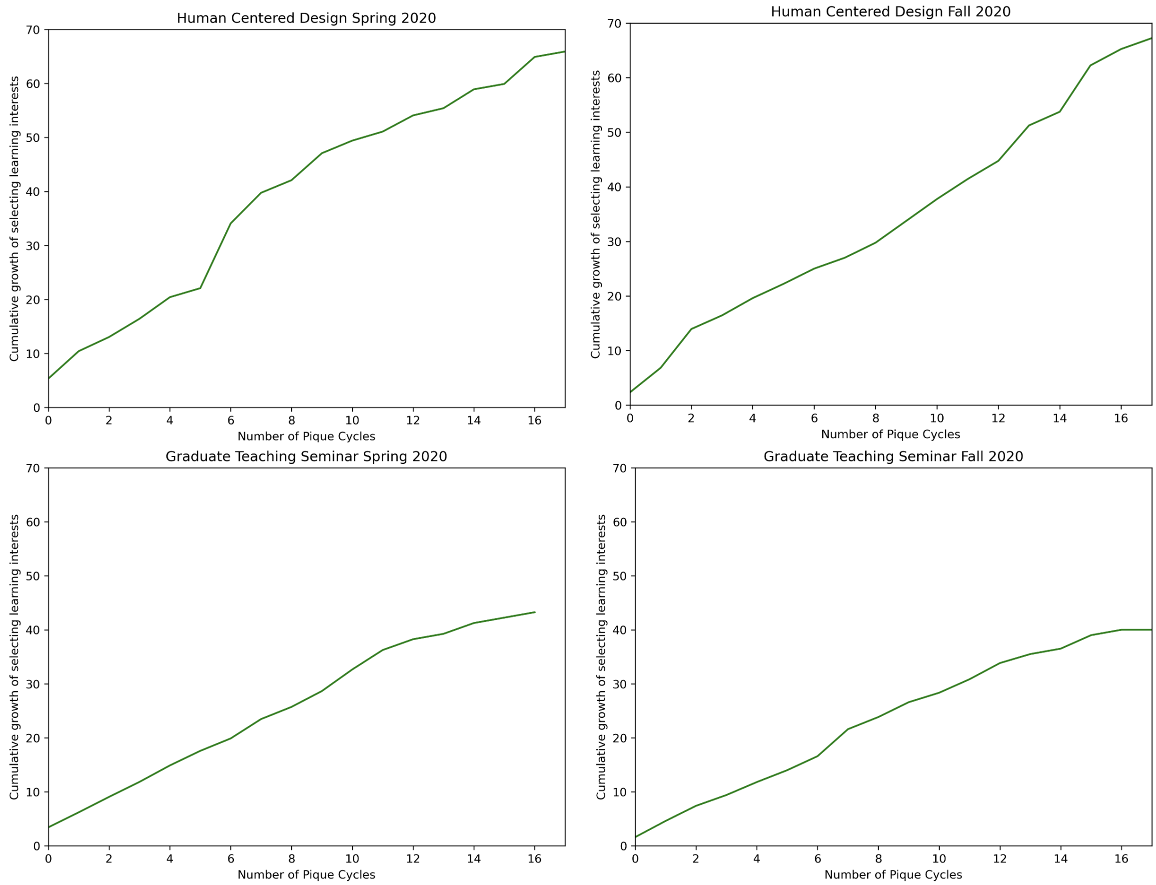



Figure 3.5: The increase in the selection of learning interests while using Pique in 2 semesters of the HCD course and 2 semesters of the Graduate Teaching Seminar [18].

As demonstrated in Figure 3.6, we found a difference between the students in the HCD courses and those in the Graduate Teaching Seminar. The top two charts are related to HCD courses in Spring and Fall semesters. The bottom two charts are related to Graduate Teaching Seminar courses in the Spring and Fall semesters. The X-axis shows the number of Pique cycles and the Y-axis shows the percentage of students searching for new interests that they had not selected in earlier cycles of using Pique. The students in the HCD courses were undergraduate and graduate students who initially expanded their learning interests and over time they reduced the number of new interests. The students in the Graduate Teaching Seminar courses were conversely PhD students who kept exploring new interests. For example, we observed that all the PhD students in the Fall semester of the Graduate Teaching Seminar continued to add new interests until the end of the semester. We observed that 71% of PhD students in the Graduate Teaching Seminar for the Spring semester had new interest in their 8th cycle of using Pique, but just 18% of the undergraduate students in the HCI course continued exploring in the 8th cycle. This result suggests that students apply the Pique system differently for expanding their learning selections.

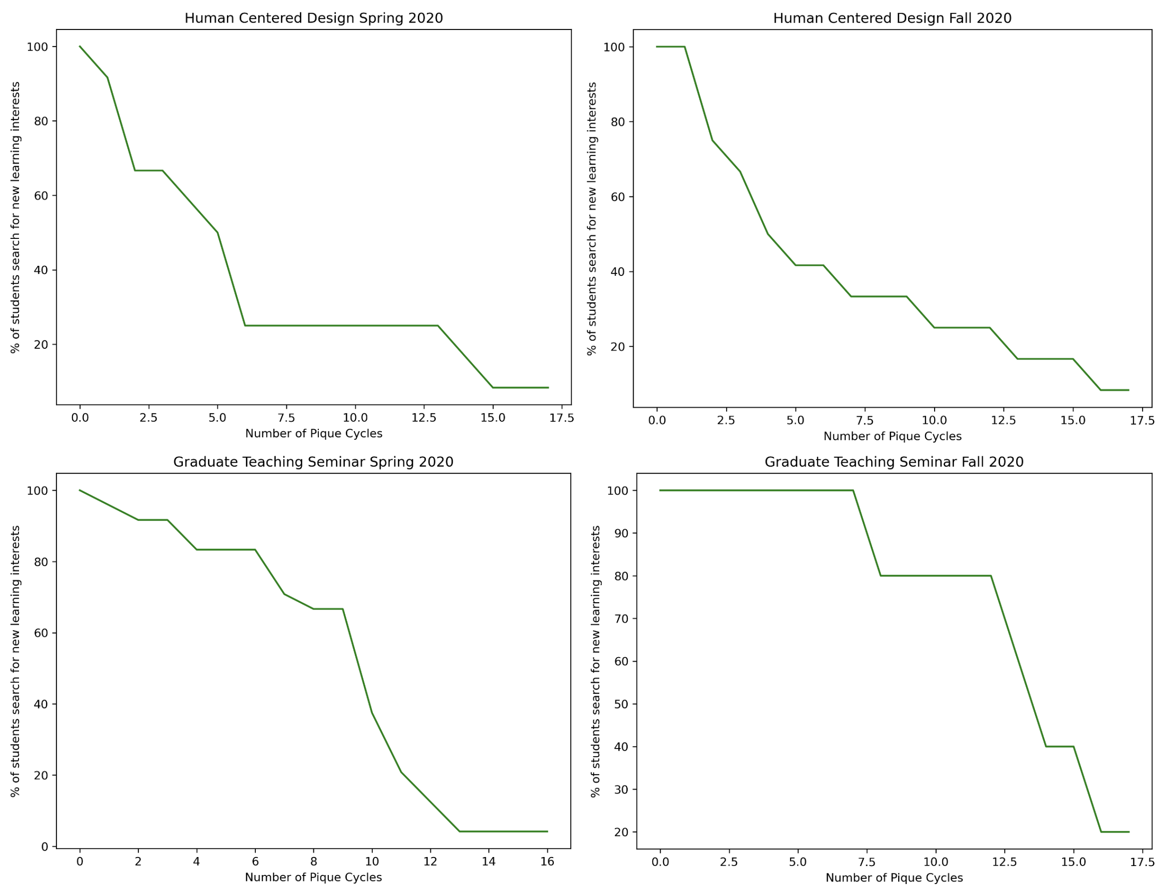



Figure 3.6: Percentage of students searching for new learning interests for each cycle while using Pique in 2 semesters of the HCD course and 2 semesters of the Graduate Teaching Seminar [18] (note that the number of cycles were different for each semester).

### 3.5.2 Qualitative Analysis of Students Experience

At the end of applying Pique by students, they were asked to reflect on which paper from the system they found most interesting and their reason for that. In order to discover meaningful patterns in the data, two researchers of our team performed a thematic analysis on students' written responses [59]. Applying multiple coders provided investigator triangulation to our analysis [60]. For having a broad consensus, initially the two researchers conducted a parallel coding workshop on the first 10% of the written responses. After discovering the initial set of themes, each of them coded the rest of the data separately and then converged on a set of collaboratively authored themes through follow-up workshops.

We discovered three major themes of novelty, personal relevance, and curiosity, underlying why most students found papers interesting. The first theme captured how students found papers interesting because of the innovation and novelty of the idea proposed in the paper. Finding novelty as one of the main themes in students' reports as their reason for why they found a paper most interesting and surprising shows that the recommended sequence of personalized and novel papers catches the students' interest and makes them surprised. The second theme captured how students found papers interesting specifically when they found its contributions and implications related to their personal life and experience. For instance, one student found a paper about a VR gaming application called "Spider Hero" interesting because they were a fan of Spiderman. Another student found a research idea of another paper so interesting because it presented new approaches for assisting disability and they had a disabled sister. Students also liked the recommended papers because they found those were aligned with their personal beliefs. For example, another student liked a paper that discussed young parents sharing information about their children online because they believed it is exactly what is happening in our society. The second theme indicates that Pique recommended personalized papers that students found interesting to read. This is consistent with the system's goal of personalization. The third theme captured how the recommended papers make students curious about the research field of HCI and computing education, and assisted to grow their interest in the field. Students get the opportunity to know about the broad research area of the field. For example, one student stated they learned something new from each of the recommended papers. They became so curious that they did extensive personal research to learn more about specific topics. We found that the curiosity theme was related to the idea of students connecting their class lessons with the recommended papers. For instance, a student learned the concept of "Wizard-of-Oz" in the HCI class sessions, and later finding the same concept in a research paper excited him a lot. The students' written responses and discovered themes indicate that the recommended papers motivated students to explore and learn more in the domain. The three main themes we identified through thematic analysis (novelty, personal relevance, and curiosity), are all consistent with the goal of Pique to recommend novel and personalized papers and consequently encourage students' curiosity to explore more and expand their knowledge.

## 3.6 Limitations

We found several limitations related to the Pique system and the data collected in the Pique. Because it was an exploratory study, these limitations are pointed out here. We had students in different classes using Pique but the study for Pique was not a controlled study. The other limitation of the study for Pique is that the number of participants was not statistically significant, and we could not do a significance test. We relied mostly on qualitative analysis and results. The other limitation is that, because we were exploring the computational models of novelty, there were different computational models of novelty in different semesters. Overall, the main limitation of the current study is the qualitative nature of the current study and the lack of quantitative results.

## 3.7 Summary

In this chapter, we addressed the second research question as, *RQ2: How computational models of novelty can be useful to encourage curiosity for students' learning* *in open-ended projects?* We showed that computational models of novelty can be applied in educational recommender systems to help in stimulating student curiosity. We presented Pique as an educational recommendation system that uses AI techniques, including NLP, to present students with personalized sequences of novel learning resources. We showed these sequences encourage curiosity and support self-directed learning. Pique applies computational models of novelty for identifying documents from a corpus of learning materials that are both relevant to the student's interest and novel with respect to the corpus. Rather than steering students through a specified curriculum, Pique aims to inspire individuals' curiosity to learn by selecting their own interests. Pique encourages students to expand their knowledge and trigger new ideas for their course projects and/or research projects by reading newly recommended learning materials.

Computational models of novelty can play a key role as a major component of the AI element in educational recommender systems for engaging learners and evoking their curiosity to explore more in the learning process. Applying an efficient novelty model in educational and recommender systems can benefit the user when accessing information by presenting the user with the most novel and surprising information among the increasingly large repositories of documents and learning materials. We applied an extension of our framework for computational models of novelty to describe the inner processes of the AI module in the Pique system. The applied framework, which is an extension of our framework demonstrated in Chapter 2, consists of four components including source data, representation method, novelty model, and personalization. This framework provides a structure for exploring and categorizing different approaches to novelty detection from the perspective of each of the 4 components in the framework and is a basis for leveraging this technology in educational recommender systems.

We developed and deployed Pique during a four-semester exploration of how to inspire students' curiosity. We chose novelty as a measure for content that encourages students' curiosity to explore more in the domain of study. The computational novelty models applied in Pique used the keywords and topics of the papers as two of the most prominent features for a scientific paper. For the first model, we used bag of keywords as the data representation method to be applied for modeling novelty. In the second model, we used Topic Modeling as the data representation method by extracting the main topics of the papers in the corpus. In the Topic Modeling approach, there is consistency in the identification of features across the entire dataset, while author-defined keywords provide features relevant to the author of a single article in the corpus.

We also developed three different models for personalization and recommendation during Pique development. The first model was just based on the student's stated interests (student's destination). The second model was based on directing them from what they already knew (student's origin) to their interests (student's destination again). The third model was based on a mixture of the origin-destination effect with similarity to the things they've recently explored. Each of these three models combined student preferences with our developed computational models of novelty to encourage curiosity in the learning process. We did not compare directly the personalization and sequence recommendation models, however, we believe, from the evidence of using them in the classroom, that both of the latter two models offer advantages over the former one.

This Chapter presents a proof of concept from the deployment of Pique, as a personalized curiosity engine and sequence generator in a recommender system for education. We have identified a number of areas for future research, as well as provide evidence of the well-known complexity and nuance of applying intelligent systems in education. We evaluated the experiences of students who used Pique as part of their courses and found three aspects that made recommended learning materials interesting: how novel they were, how personally relevant they were, and the curiosity and further self-directed learning that they evoked. Our findings are evidence of how curiosity can be elicited from students as part of a course experience when self-directed and open-ended engagement with learning resources is desirable. Our results from reflection surveys and written reports indicate that students were interested in the personalized papers recommended. We observed that students are eager to engage with educational recommender systems like Pique and that their interests diversified as a result. While this study is limited by its lack of a control (educational controls are notoriously challenging both due to the difficulty of controlling for all possible confounds as well as the moral dubiousness of withholding the hypothesized "best" instruction from some students), it does show the promise of curiosity-driven recommendation. Developing educational systems like Pique can help students expand their knowledge by recommending novel scientific articles. While we cannot claim that student curiosity was entirely due to Pique, we conclude that the approach of encouraging curiosity Pique shows is promising for our future research on computational novelty in open-ended learning environments.

We showed 2 major contributions in this chapter: a framework for structuring the AI component of educational recommender systems to encourage curiosity and the Pique model that integrates the AI component and its interaction with a learner model and course materials. We demonstrated how the framework is integrated in the Pique model, providing opportunities for future studies that leverage other models of novelty and personalization. The students' experience with Pique was described demonstrating how their interests expanded over the period of a semester. Future studies that collect data from a larger number of students would allow an analysis of the relationships between students' expanding interests and the novelty score of the recommended learning materials.

In this chapter, we described the role and usefulness of computational models of novelty in educational recommender systems to encouraging students' curiosity. Pique as a primary study showed how the computational novelty framework can be applied as the basis for the AI in educational recommender systems. In the next chapter, we emphasize using Topic Modeling as the data representation method in building computational models of novelty and introduce two novelty models based on Topic Modeling. We explain how Topic Models can be the basis for and can facilitate measuring novelty of research publications. We then compare our computational models of novelty based on Topic Models with human perception of novelty by running a study and recruiting experts in the domain of our dataset (HCI) and report on the results.

# CHAPTER 4: TOPIC CO-OCCURRENCE VS SIMILARITY AS COMPUTATIONAL MODELS OF NOVELTY USING TOPIC MODELING

Chapter 4 addresses research questions 3, 4, and 5 respectively:

RQ3: How can computational novelty models of research publications be defined by using Topic Modeling, based on topic combination and similarity approaches?

RQ4: How do topic combination and topic similarity measure novelty differently on the same corpus of research papers? How does the novelty score distribution differ? And, how is the meaning of novelty expressed differently in the two models: co-occurrence (combination) and similarity of topics?

RQ5: How do computational models of novelty compare to human perception of novelty?

We developed two computational models of novelty using Topic Modeling, one based on atypical combination of topics (topic co-occurrence), and the second based on similarity of topics. Despite the approaches applying author-defined keywords in data representation, applying Topic Modeling provides consistency in the identification of features across the entire dataset, but approaches like author-defined keywords provide features relevant to the author of a single article in the corpus which lacks consistency. In this chapter, we show the process of developing our computational models of novelty from the perspective of the three components of our suggested framework in chapter 2 (Figure 2.1): source data, representation method, and novelty models. Our study fits into the framework as follows: our source data type is research publications (described in section 4.1), our data representation method is Topic Modeling (discussed in section 4.2), and our described novelty models are based on atypical combinations and similarity of topics (discussed in sections 4.3 and 4.4). We demonstrate the results and analysis of applying Topic Modeling and computational novelty models on our dataset of HCI research publication abstracts. The degree of novelty of a paper is determined by applying two different approaches:

topic combination approach, by the degree of atypicality of the topic combinations in the paper, and similarity approach, by the degree of distance between the vector of a paper and the corpus average. We then compare our computational models of novelty based on Topic Models with human perception of novelty by running a study and recruiting experts in the domain of our dataset (HCI) and report on the results (section 4.6).

## 4.1 Dataset (Source Data)

We collected a total of 1974 research paper abstracts from ACM CHI (ACM Conference on Human Factors in Computing Systems) proceedings published between the years of 2017 and 2019. ACM CHI is one of the most influential conferences in the field of human-computer interaction. Our dataset contains 604 papers published in the year 2017, 667 papers published in 2018, and 703 papers published in 2019. We extracted the papers' abstract, title, authors, URL address, and issue year. In our research we used the abstract of the papers as a summary of each paper for running Topic Modeling and novelty models on them. The paper title and publication year are used for our further explanation and investigation purposes. In the next section, we describe Topic Modeling as the data representation method for text data (the second component of the framework in Figure 2.1). In the following sections, we describe data preprocessing, results of running Topic Modeling on the dataset, computing the novelty of research paper abstracts based on the two novelty models with results, and report on the study we run to compare the computational models of novelty with human experts perception of novelty.

## 4.2 Topic Modeling in Representing Unstructured Text Documents

The approaches for modeling novelty are primarily dependent on how we are representing the data. As shown in the framework in section 2, the novelty model is built on the data representation. Appropriate representation of data plays an important role in achieving an effective novelty measurement. Among several other representation methods such as bag of words, TF-IDF, word embedding, etc. Topic Modeling is a less examined approach to be applied in modeling novelty of text data, in particular research publications, which is our focus in this dissertation. In this section, we describe Topic Modeling and its most well-known algorithms as one of the data representation methods for text documents.

### 4.2.1 Definition of Topic Modeling

In natural language processing and machine learning a Topic Model is a type of statistical model to discover the main topics of documents in a corpus. Each extracted topic consists of a probability distribution over words and each document consists of a probability distribution over topics [1, 31]. Depending on the specific topic(s) of a document, particular words appear in the document more or less often. For example, "galaxy" and "comet" will occur more in documents about astronomy, and "medicine" and "hygiene" will occur more in documents about health. Each document in the corpus is assigned with different proportions of each topic and a document may discuss several topics in different proportions. The topics discovered by Topic Modeling techniques are groups of similar words which may not always consist of a single subject that is easily human understandable, but they often can be interpreted in some way.

Topic Models are also introduced as probabilistic Topic Models, referring to statistical algorithms and text mining techniques for finding the hidden semantic structures in the corpus of text documents. The amount of text information we can easily access in recent decades is beyond the processing capacity of humans. Topic Models can help to categorize and provide reliable insights for us to understand large collections of unstructured text bodies [1,61]. Figure 4.1 (adapted from Blei, Probabilistic Topic Model [1]), shows the overall concept of Topic Model.

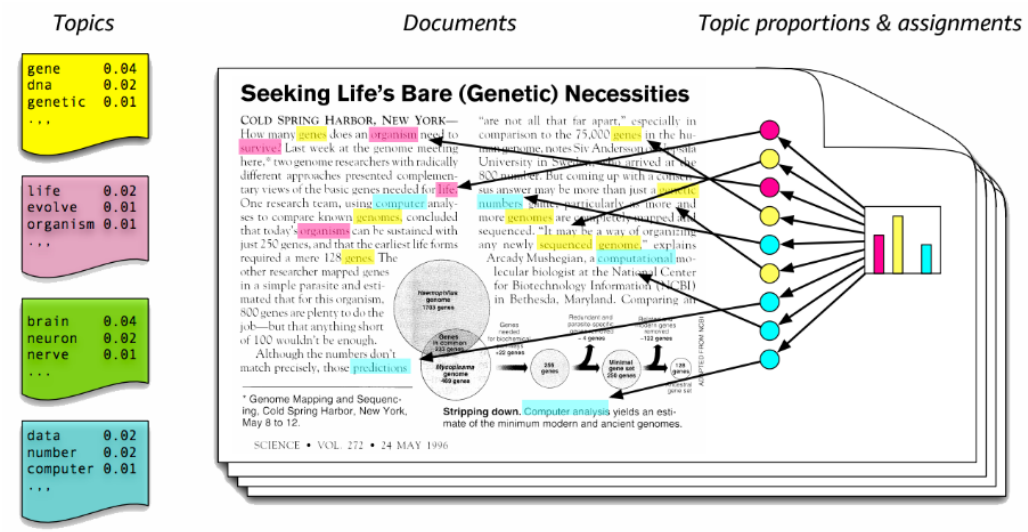



Figure 4.1: Topic Modeling to discover hidden semantic structures in a corpus [1].

Each document is initially considered as a bag of words ignoring word order and context in order to provide a unified vector representation for each document [48]. Indeed, each document is represented as a vector with length equal to the vocabulary size for the corpus while each dimension of this vector corresponds to the frequency of a word in the document [1, 31, 48, 52]. Then ultimately the Topic Model delivers two separate probability distributions: distribution of words for each topic, and distribution of topics in each document. Topic Models may deliver more output depending on the type of the Topic Model algorithm. In the next section, we review three of the most well-known Topic Model algorithms and demonstrate the model we are using in our research for representing scientific papers.

### 4.2.2 Topic Model Algorithms

Several probabilistic Topic Model algorithms have been introduced for representing text documents each having specific features. One of the most basic Topic Modeling algorithms is Latent Dirichlet Allocation (LDA) introduced by Blei et al. [31]. LDA applies Dirichlet distributions for building topics per document model and words per Topic Model (see [31] for the full theory behind LDA). Figure 4.2 (adapted from [62]) illustrates the LDA model visually for a sample corpus of text documents. It also illustrates the basic model for several other Topic Model algorithms discussed later. The goal of the model is to find the topic and document probability distributions (vectors) which explain the observed data that is the original bag of word representation of the different documents [62]. Each of the documents will be represented by a vector of length $K$ including proportions between 0 and 1 that describe which topics exist in that document; where K is the number of topics initially set for training the Topic Model. The document vectors are mostly sparse, with low dimension and high interpretability, demonstrating the pattern and structure in documents. A document may consist of say 60% Topic 1 and 30% Topic 2. The model often results in document vectors with many zeros indicating there are a few number of topics appearing per document. That is, documents typically only discuss a limited number of topics which increases human interpretability of these document vectors [62]. Each of the K topics is represented by a vector of length V that describes which words are probable to occur, given a document on that topic (assuming the vocabulary in the documents consists of V words). For example for topic 1 in the sample model shown in Figure 4.2, "school", "student" and "learning" could be some of the most common words [62]. This can be interpreted as the "Education" topic. For topic 2, the words "director", "cinema" and "actor" might be the most common words which can be interpreted as the "movie" topic.

CTM (Correlated Topic Model) is another Topic Model algorithm introduced by Blei et al. [52] which is an extension of the basic Topic Modeling algorithm. The advantage of Correlated Topic Models over the LDA is that they capture the correlation between the topics. For example, a topic about genetics appears more frequently with a topic about disease than it does with a topic about astronomy because genetics and disease themes are more conceptually related [48]. A weakness of LDA is its incapacity to model topic correlation that is because of the use of the Dirichlet distribution to model the variability among the topic proportions [52]. In CTM the topic proportions exhibit correlation via the logistic normal distribution [63] (more details can be found in [31, 52, 64]). Blei et al. [52] apply the CTM on a dataset of the articles from Science journal comprising 57M words and discuss that "*CTM gives a better fit of the data than LDA and can be used as an effective exploratory tool for better understanding a large corpus of document*" [52].

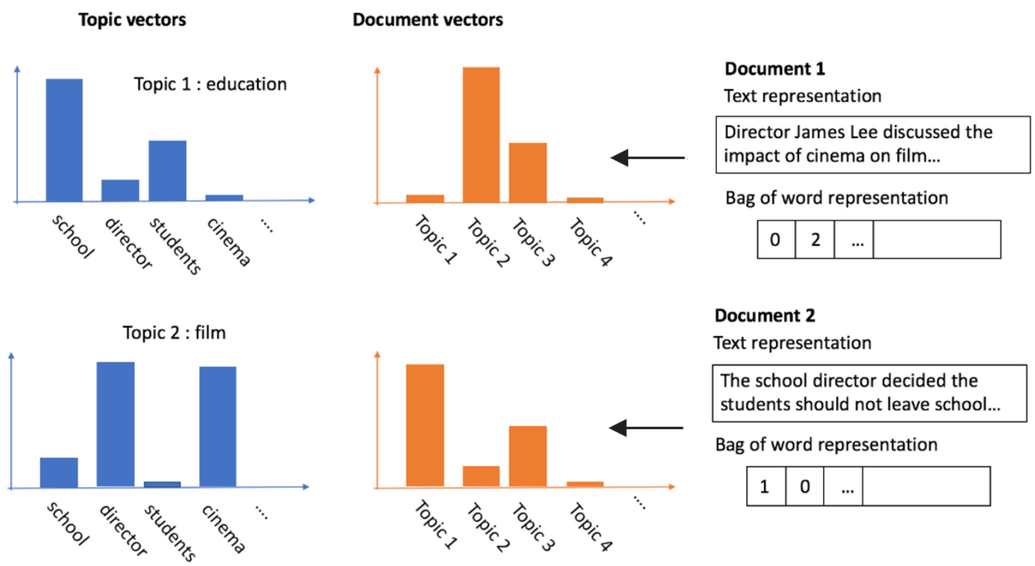



Figure 4.2: Topic and document vector representation in LDA model for a sample corpus of text documents [62].

STM (Structural Topic Model) [42] is another Topic Model extension and with some configurations is similar to CTM implementation. STM includes document-level metadata into the standard Topic Model (see [65, 66] for more details). Similar to CTM, STM topic model is not based on the preliminary assumption made by basic Topic Model algorithms like LDA which assume: all the topics in a corpus are independent and thus no pair of topics is more likely to occur together in a document than the others. STM considers correlation between topics as CTM does. All the discussed Topic Model algorithms deliver topic vector and document vector respectively for all topics and documents of the corpus. CTM and STM further deliver correlation for all topic pairs.

Considering the advantage of capturing the correlation between topics, and the comprehensible and appropriate functions STM provides, in our study, we applied STM (R package "STM" [65]) for representing research publications in order to build our novelty models on top of that. Figure 4.3 demonstrates our data representation model using STM. As shown in Figure 4.3, STM runs on a corpus of text documents where $V$ is the number of vocabularies appearing in documents and $K$ is the STM parameter for the number of topics that should be set initially. It is found that lower number of topics delivers more semantically distinguished topics, however, higher number of topics may give better fit models but reduce semantic coherence of topics [8,67]. STM provides a useful function called *searchK* which gives the best value of $K$ among different input values in terms of semantic coherence. We applied the *searchK* function and tested 10, 20, 40, and 100 as different numbers of $K$ (not very high numbers to prevent losing distinguishability of topics resulting from a large $K$) [67]. In our study, we used 20 as $K$ (which is also the default number of topics in STM) giving the most coherent and discernible topics identified both by the result of the *searchK* function and also by our observation and interpretation of the output topics.

STM delivers 3 main outputs: topic vectors, document vectors, and a correlation matrix. Each topic is represented with a $V$ dimensional vector containing the distribution of words in that topic. Each document in the corpus is represented with a $K$ dimensional vector containing the distribution of topics in that document. The correlation matrix is a $K \times K$ matrix where each element contains the correlation coefficient of a pair of topics indexed by row and column number. We demonstrate the result of running STM topic modeling on our dataset of research publication abstracts in the next section.

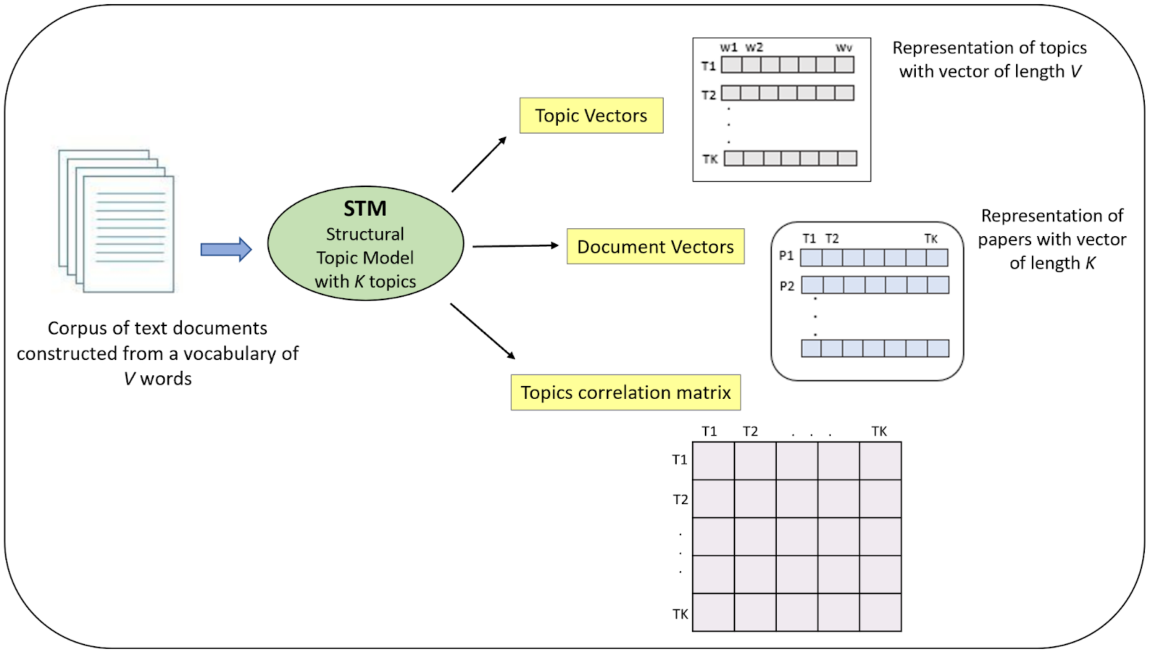



Figure 4.3: Finding topics in scientific publications with STM.

In our research, we applied Topic Modeling for representing documents in our dataset since it automatically provides consistent topics by scanning and considering all the dataset consistently, not just by some human expert opinion (e.g. first approach in [16] in which a human expert assigns topic labels to the health news documents, or [26] in which the author of each paper assigns keywords to a single paper). By applying Topic Modeling we consider each paper as a point in vector space. Representing papers as a vector with elements as numbers between 0 and 1 facilitates novelty modeling based on different approaches including similarity and topic combinations. In the next section, we show the results of running topic modeling in our dataset of research publication abstracts. Then we describe two computational novelty models built by applying Topic Modeling in sections 4.3 and 4.4, one based on topic combination and another based on similarity. We demonstrate the application and result of two novelty models on our dataset.

### 4.2.3 Results of Topic Modeling on the dataset of HCI papers

We used the R package "STM" [65] to construct the novelty models. In order to prepare the data to feed to the STM algorithm, we removed the stopwords, numbers, and punctuations. All the words were converted to lowercase and then stemming was performed. After preprocessing steps, we run STM topic model algorithm on our dataset to extract the main topics of the corpus of paper abstracts. After testing different values for the number of topics (10, 20, 40, 100) and comparing the results and coherence of the topics, we used 20 (the default number of topics in STM algorithm) in our study giving the most distinguishable and coherent topics. As shown in Figure 4.3, the three main outputs we obtained from the Topic Model on top of which we build our novelty model are: 1)Topics derived by STM algorithm, 2)document vectors including proportion of topics for each document, and 3)Topics correlation matrix. Each document in the corpus is represented with a 20 dimensional vector containing the distribution of topics in that document. For our study, the correlation matrix is a $20 \times 20$ matrix reflecting 20 topics of our model and containing the correlation coefficient for all pairs of topics (see Figure 4.3).

To explore the words associated with each topic, we used a function in the STM package called *labelTopics* [65]. The function prints four different types of word profiles for each topic, including *highest probability* words, *lift* words, *FREX* words, and *score* words [65]. Each of these names refers to a different formula for displaying the most representative words in each topic (for more information on *high probability*, *FREX*, *score*, and *lift* see [42, 68–71]). We investigated these four lists of words for all 20 topics to understand the subject of topics. In order to better demonstrate the distinction of the subject of topics, for each topic we extracted the top most connected (conceptually related) and informative words among the lists of word profiles. Table 4.1 shows the top most representative words for each of the 20 topics. Some of the words seem not to be the complete word as they are the stem of the combination of several words with the same root. Most of the obtained topics have understandable meanings in the domain of HCI research. Our interpretation of the subject of each topic is included in the last column of Table 4.1.

Table 4.1: Top most representative words for 20 topics and our interpreted subject for topics.

| Begin of Table | | |
|---|---|---|
| Topic # | Most representative words | Our interpretation |
| 1 | *privaci, control, algorithm, person, trust, inform, user, drone, risk, explan, safeti, concern, secur, polic, bureaucrat, cyberbulli, leakag* | Privacy control, trust, safety |
| 2 | *communiti, home, practic, smart, citizen, citi, urban, thing, bed, technolog, hci, animal-comput, authorit, grassroot, placemak* | Smart home/life, smart city |
| 3 | *visual, studi, data, user, peopl, impair, access, chart, blind, reader, map, color, magnif, multiclass, referenc, user-driven, behalf* | Visualization, Accessibility |
| 4 | *design, data, interact, paper, hci, process, approach, reflect, research, workshop, concept, framework, fiction, ethic, hackathon, aborigin, co-occurr, coffe, conceal, disciplinari, first-hand* | Interaction design, ethic |
| 5 | *social, share, media, peopl, particip, experi, support, dementia, photo, facebook, persona, news, activist, broadcast, folk, larp* | Social media |
| 6 | *health, patient, support, technolog, design, person, care, clinic, caregiv, self-track, diseas, clinician, suicid, emoji, self-car, bipolar, cessat, clue, counsel* | Health |

| Continuation of Table 4.1 | | |
|---|---|---|
| Topic # | Most representative words | Our interpretation |
| 7 | *model, user, predict, data, task, human, system, chatbot, agent, label, dataset, comparison, pairwis, accuracy, saccad, coder, ecg* | Predictive models |
| 8 | *user, app, mobil, studi, particip, smartphon, data, interrupt, password, authent, file, attack, secur, email, reset, mturk, defens, encrypt* | Smart phone, security, mobile app |
| 9 | *digit, work, support, report, technolog, inform, challeng , infrastructur, stakehold, blockchain, transact, workplac, mind, worker, employe, cash, chariti, corrupt* | Digital work, digital currency |
| 10 | *game, player, experi, play, studi, driver, design, drive, vehicl, car, transport, gameplay, uber, commut, drew, extrins, finland, gamer, hone* | Interactive games, car apps |
| 11 | *collabor, feedback, task, work, worker, provid, team, crowd, plan, intellig, crowdsource, behavior-chang, self-awar, checklist, communal, microtask, planner* | Collaboration, crowdsourcing |
| 12 | *communiti, onlin, particip, gender, search, inform, women, group, older, rumor, find, websit, job, older, mentorship, buy, catalog* | Gender |
| 13 | *user, interact, effect, design, link, warn, phish, taxonomi, advertis, interface, method, inform, studi, medit, url, email, chrome, cross-devic, habitu, aid* | Web security, user interfaces |

| Continuation of Table 4.1 | | |
|---|---|---|
| Topic # | Most representative words | Our interpretation |
| 14 | *tool, video, user, system, algorithm, code, program, develop, sketch, novic, api, callback, choreograph, metacognit, traine, higher-level* | Video tools, algorithm and programming, sketching |
| 15 | *studi, cue, particip, user, robot, technolog, auditori, sound, stimuli, speech, audio, biofeedback, poke, tactil, vibrat* | User study, technology, speech, audio |
| 16 | *gestur, devic, input, touch, interact, user, technique, finger, touchscreen, latenc, target, accuraci, bod, chassi, coeffici, corner* | Touch/gesture interaction |
| 17 | *learn, children, student, educ, parent, design, support, teacher, school, learner, classroom, classmat, classroom, curricular, detector, faculti, infant* | Children, Education |
| 18 | *virtual, user, realiti, object, physic, feedback, haptic, immers, haptic, augment, render, manipul, object, actuat, teleport, airflow, congruent, disorient, video-medi* | Virtual reality, haptic |
| 19 | *design, interact, print, fabric, textil, materi, circuit, properti, object, prototype, electron, stretchabl, textur, batteryless, bend, cad, chip, connector* | Interactive textiles |

| Continuation of Table 4.1 | | |
|---|---|---|
| Topic # | Most representative words | Our interpretation |
| 20 | *text, method, type, user, error, keyboard, studi, keyboard, entri, word, wpm (words per minute), speed, dwell, nonvisu, paragraph, sentenc, tilt-bas, alphabet* | Text interfaces |
| End of Table | | |

For example, Topic 1 seems to be mostly about privacy control, trust, and safety. Topic 2 is about smart homes and smart cities. Topic 3 is clearly about visualization and accessibility, and Topic 5 is about social media. Topic 6 is related to a particular research subject, that is health. Our interpretation of other topics can be found in Table 4.1. We observed some meaningless words in some topics like "wpm" in Topic 20. We searched our dataset for these kinds of meaningless words and found that they are abbreviations for some phrases; that is "*words per minute*" for "wpm". So we included the actual phrases in parentheses next to the abbreviations in Table 4.1 for clarification. This model reasonably reflects the abstracts of our dataset with most of the topics having clear meaning in the HCI field.

As stated earlier, each cell in the correlation matrix contains the correlation coefficient for a particular topic pair corresponding to the row and column numbers of the cell. Topic Modeling identified Topic 4 (ethics, interaction design) and Topic 16 (touch/gesture interaction) as the most atypical topic combination in our dataset of abstracts with the most negative correlation coefficient in the correlation matrix. That means in our dataset, these two topics do not appear/co-occur together in a paper as much as other topic pairs do (i.e. they are considered as an uncommon topic combination by our model). However, both of these topics seem to have some overlap in interaction related subjects but we think that is because of the nature of our dataset which is Human-Computer Interaction, so all the papers are related to HCI matters somehow. Regardless of our (human) interpretation, the underlying logic of Topic Modeling found these two topics as the least co-occurring ones.

Topic 2 (smart home, smart city) and Topic 4 (ethics, interaction design) have the most positive correlation coefficient, identified by the Topic Model as the most correlated topic pair in the corpus. The model recognized that these two topics co-occur together in a paper more than other topic pairs in our dataset of abstracts. That means many papers in our corpus include these two subjects.

## 4.3 Atypical Combination of Topics for Modeling Novelty

One approach for identifying novel items regardless of the domain is looking for items with atypical (unexpected or uncommon) conceptual combinations. Novelty can be defined as any new combination of items that is not similar to the past frequent observed combinations [12, 25, 48, 72]. This also can be explained as observing any combination of items with low probability of co-occurrence which we call atypical combination.

In this section, we explore how the atypicality of topic combination can be applied for modeling novelty of research publications. The strategy for doing so is to use advanced Topic Model algorithms that provide us with correlation between topics for representing the text documents and then identifying the papers with the most uncommon topic combinations as the most novel papers in the corpus. We presume applying Topic Modeling can facilitate computational novelty discoveries. This section describes an approach for calculating the novelty of papers in a corpus of research publication abstracts considering the topic combinations in the paper abstracts. We first review the process of identifying concepts (topics) within text documents and determine their relationships. Then we assess the atypicality (novelty) of the combinations of the most representative topics that appear in abstracts of the scientific papers.

We base our models of novelty on STM Topic Modeling algorithm to consider the correlation between topics. Topics can be more or less correlated. For example, a topic about medicine occurs more often with a topic related to health than with a topic about Art because those topics are more relevant conceptually. This constructs the basis of our novelty model presuming "*topics are concepts concluded from the dataset, and the correlations between topics give us a basis for what combinations of concepts are unexpected (novel)*" [48]. We compute the overall novelty for each paper/document by considering the pairwise correlation of the top 5 topics with highest topic proportions within the document. Equation 4.1 shows the novelty measurement formula for paper abstract $P$ given $P = [t_i, t_j, \ldots, t_n]$ consisting of the set of top $n$ presented topics in $P$. In our research, we set $n$ equal to 5 (considering the top 5 topics of each paper as its most representative topics) but it could be set to any number less than $K$ (the number of topics extracted by Topic Model that is 20 in our study).

$$Novelty\ P = \sum_{i=1}^{n} \sum_{j=i}^{n} \left(\frac{CovMat(t_i, t_j)}{CovMat(t_a, t_f)}\right) \times (min(prop(P, t_i), prop(P, t_j))) \tag{4.1}$$

$CovMat$ is the covariance matrix obtained from the Topic Model [48]. $CovMat(t_i, t_j)$ is the correlation of the topic combination $(t_i, t_j)$, and $CovMat(t_a, t_f)$ is the correlation of the most atypical topic combination among the whole corpus (that is used for normalization). $t_a$ and $t_f$ are the pair with the least correlation in the $CovMat$ matrix. $prop(P, t)$ is the proportion of document $P$ that consists of topic $t$. The expression in the first parentheses calculates the novelty for each of the topic combinations (pairs) of the document's top $n$ topics, represented as a proportion of the most novel topic pair ($t_a$ and $t_f$ ) in the model. The number $n$ should be between 1 and $K$. In our research, we set $n$ to 5 considering the top 5 topics of the paper as its most representative topics. We found that the remaining topics for most of the papers have very low proportions not considerable for novelty calculation. The second expression of the formula is the minimum of the two proportions of document $P$ related to topics $t_i$ and $t_j$. The product of the two expressions provides the normalized novelty for each topic pair weighted by how much of the $P$ consists of that combination [48]. The overall novelty score for each paper is then calculated as the sum of this product for all of the possible topic combinations obtained from 5 top topics in $P$. We decided to consider the correlation and combination for all possible topic pairs from the top topics in the paper abstract in the novelty calculation, rather than just considering its most atypical topic pair.

The reason for using the minimum of the two topic proportions of $t_i$ and $t_j$ rather than the sum of them is to prevent favoring documents that do not have a significant proportion with one topic, and are not especially surprising. Therefore, the minimum of the topic proportions for these two topics is used in the formula to weight the novelty measure towards papers containing significant amounts of both topics [48]. Figure 4.4 demonstrates the overall novelty model using topic combination approach. Papers consisting of only topics with negative correlation will receive a positive novelty score and vice versa. The novelty score for papers with a lot of relatively unexpected/novel topic combinations will be higher than the score for documents containing only a small portion of very novel topic pair.

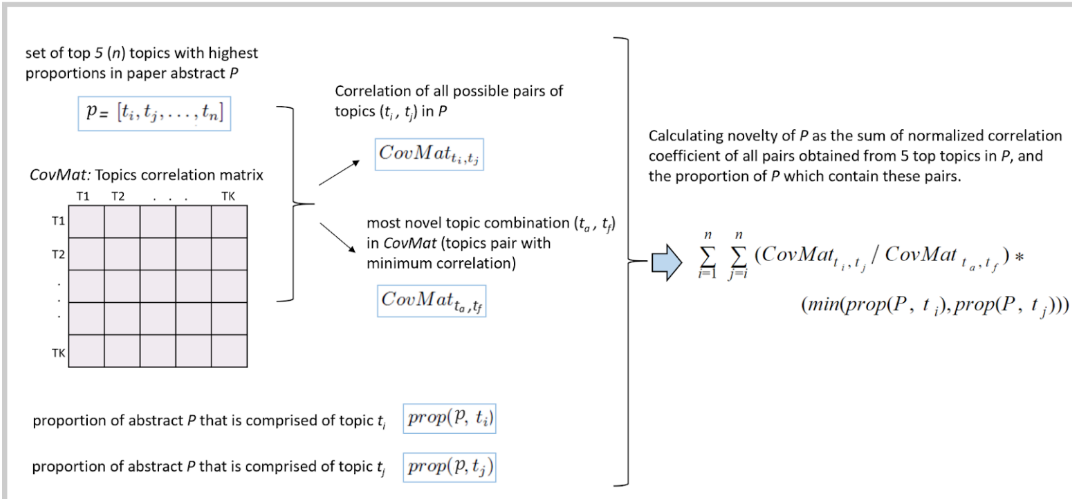
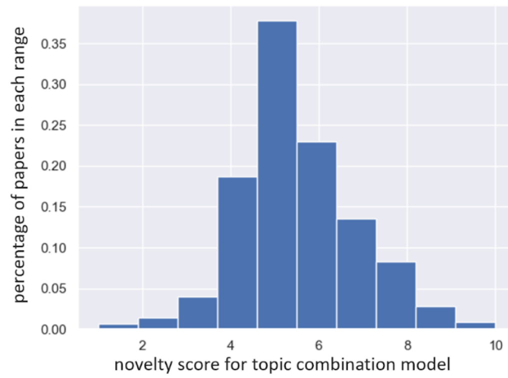



Figure 4.4: Novelty model based on atypical combination of topics.

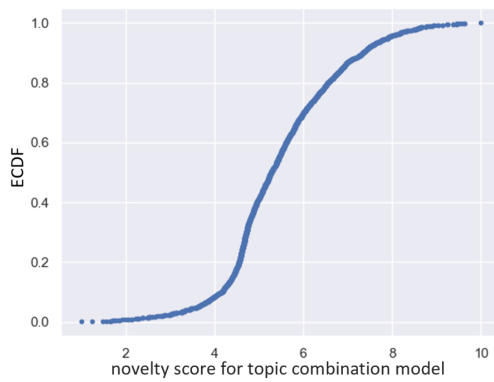
Applying advanced Topic Model algorithms (that capture the topic correlation) for data representation facilitate computational novelty discoveries of research publications by considering the atypicality of topic combinations in a paper. In this section, we described an approach and methodology for modeling the novelty of scientific papers using Topic Modeling and considering the pairwise combinations of the 5 most representative topics of the paper. In the next section, we describe applying this novelty approach with examples of applying and the results on our dataset.

### 4.3.1 Application and Results for Novelty as Atypical Combination of Topics in HCI Papers

Using the output from the Topic Model, we developed our first novelty model based on topic combination on our dataset. We assigned a novelty score to each paper based on this model (see equation 4.1) and ranked the papers by their novelty scores. Figure 4.5 shows the distribution of novelty scores based on the topic combination approach. We can see that the distribution for this model is rather close to a normal distribution. The probability density function (PDF) and cumulative distribution function (CDF) of novelty scores are illustrated in parts (a) and (b) of Figure 4.5 respectively.




(a) Probability density function for topic combination model




(b) Cumulative distribution function of novelty scores

Figure 4.5: Distribution of novelty scores based on the topic combination approach.

We exhibit three of the most novel papers, three from the moderate novel ones, and three of the least novel papers in our dataset as examples. Tables 4.2, 4.3, and 4.4 show respectively these three sets of papers within our dataset. For picking the three examples of moderate novel papers, we randomly selected three papers within the middle 40% of the sorted list of papers based on their novelty score. For each paper listed in the three tables, we also display the novelty score as well as the most prominent/representative topic pair involved in the computation of novelty score for that paper. By "most prominent pair", we mean the pair with the highest weight/value (for the multiplication/product of correlation coefficient and the topic proportion) in novelty calculation among all possible topic pairs obtained from the top 5 topics of the paper (see equation 4.1).

Table 4.2: Three most novel paper abstracts in the corpus for the first novelty model.

| **Title** | **Novelty score** | **Most prominent pair** |
|---|---|---|
| 1: "Keppi: A Tangible User Interface for Self-Reporting Pain" (2018) | 10 | 6 (health),16 (touch/gesture interaction) |
| 2: "Enabling Identification and Behavioral Sensing in Homes using Radio Reflections" (2019) | 9.64 | 2 (smart home, smart city),14 (video tools, algorithm and programming, sketching) |
| 3: "SmartManikin: Virtual Humans with Agency for Design Tools" (2019) | 9.60 | 4 (interaction design and ethic),18 (virtual reality, haptic) |

As shown in Table 4.2, the first novel paper identified by the topic combination novelty model talks about a new pressure-based tangible user interface for self-reporting pain [73]. This paper incorporates topic 6 (health) and topic 16 (touch/gesture interaction), two typically not related topics. The novelty model picked up this abstract as the most novel one. Figure 4.6 displays the abstract for this paper as an example to demonstrate how the algorithm picks/identifies the most novel paper in the corpus by applying topic combination as the novelty model on top of Topic Modeling as the representation method. The highlighted parts demonstrate clear examples of the combination of topic 6 and topic 16 captured by the Topic Model and novelty algorithm.

> *"Motivated by the need to support those managing chronic pain, we report on the iterative design, development, and evaluation of Keppi, a novel pressure-based tangible user interface (TUI) for the self-report of pain intensity. In-lab studies with 28 participants found individuals were able to use Keppi to reliably report low, medium, and high pain as well as map squeeze pressure to pain level. Based on insights from these evaluations, we ultimately created a wearable version of Keppi with multiple form factors, including a necklace, bracelet, and keychain. Interviews indicated high receptivity to the wearable design, which satisfied additional user-identified needs (e.g., discreet and convenient) and highlighted key directions for the continued refinement of tangible devices for pain assessment."*

Figure 4.6: Abstract of the most novel paper selected by topic combination novelty model.

The second most novel paper picked by the novelty model suggests a system that automatically gathers behavior data at home without any sort of user input or wearing sensors and by transmitting wireless signals and analyzing its reflection from the environment [74]. This paper incorporates topic 2, the smart home topic, with topic 14, which is associated with programming and algorithms. The paper describes novel algorithms for recognizing "*who does what*" at home and bootstrapping the system in new homes without requiring users for new annotations [54]. Topic Model detected a negative correlation coefficient for topics 2 and 14.

The third most novel paper introduces a virtual human called SmartManikin to improve designing comfort, interaction, and usability in products [75]. This paper incorporates topic 18 (virtual reality, haptic) with topic 4 (interaction design, ethics). Topic Model recognized a negative correlation coefficient between these two topics. The paper describes that *SmartManikin* improves design by providing real-time feedback on design changes regarding comfort and usability [75].

Table 4.3: Three of the moderate novel paper abstracts in the corpus.

| **Title** | **Novelty score** | **Most prominent pair** |
|---|---|---|
| 1: "Communicating Algorithmic Process in Online Behavioral Advertising" (2018) | 5.45 | 1 (privacy control, trust, safety), 4 (interaction design and ethic) |
| 2: "How Design-inclusive UXR Influenced the Integration of Project Activities: Three Design Cases from Industry" (2017) | 5.28 | 4 (interaction design and ethic), 11 (collaboration, crowdsourcing) |
| 3: "Exploring and Designing for Memory Impairments in Depression" (2019) | 4.90 | 6 (health), 15 (user study, technology, speech, audio) |

The first paper in table 4.3 incorporates topic 1, which is associated with safety, privacy control, and trust, with topic 4, interaction design and ethics. This paper discusses how personal advertisement algorithms may violate user privacy and consequently reduce user trust and desire in behavioral advertising [76]. The second moderate novel paper investigates how the rearrangement of project activities to consider design-inclusive User Experience Research (in which design is an essential part of research) can advance UX design qualities compared to the complete separation of design activities and UXR [77]. This paper incorporates topic 4 (interaction design and ethics) with topic 11 (collaboration, crowdsourcing) as its most representative topic pair. The third paper in Table 4.3 incorporates topic 6 which is related to health with topic 15 (user study, technology, speech, audio). This paper suggests considering depression related memory impairment within the design to help for treatment of memory disorders caused by depression [78]. It reports on interviews with experts in treating depression and discusses new design opportunities for memory technologies for depression [78].

Table 4.4: Three of the moderate novel paper abstracts in the corpus.

| **Title** | **Novelty score** | **Most prominent pair** |
| --- | --- | --- |
| 1: "Creating a Sociotechnical API: Designing City-Scale Community Engagement" (2017) | 1 | 2 (smart home/life, smart city), 4 (interaction design, ethic) |
| 2: "Technology and the Givens of Existence: Toward an Existential Inquiry Framework in HCI Research" (2018) | 1.25 | 2 (smart home/life, smart city), 4 (interaction design, ethic) |
| 3: "Deployments of the table-non-table: A Reflection on the Relation Between Theory and Things in the Practice of Design Research" (2018) | 1.50 | 2 (smart home/life, smart city), 4 (interaction design, ethic) |

All three low novel papers displayed in Table 4.4 incorporate topic 2 with topic 4, the most conventional topic pair in the corpus identified by Topic Model (having the highest correlation coefficient). Topic 2 is about smart home/life and smart city, and topic 4 is associated with interaction design and ethics. The abstract of all three papers are kind of describing building on previous attempts and other literature [79–81]. By reading them, we found their research is actually complementing current approaches in HCI and reviewing related works. Novelty model recognized these as the least novel kind of paper in the corpus.

We observed that for all 3 papers in Table 4.2, the correlation coefficient of the most prominent pair is negative, signifying that higher atypicality of topic combination gives a higher novelty score. However, for all the three least novel papers listed in Table 4.4, the correlation coefficient of this pair is positive, meaning that the papers captured by the model as the least novel ones consist of conventional topic combinations. We found that the most atypical topic pair in the dataset has a role in calculating the novelty score for one of the three most novel papers (the third novel paper has topics 4 and 16 among its top 5 topics). In contrast, none of the 3 least novel papers have the combination of topics 4 and 16 (most atypical pair) among their 5 highest topic proportions. Indeed, this uncommon topic combination has no role in calculating the novelty score of the least novel papers. We found a small correlation of 0.02 between the novelty score and year of publication in the first novelty model based on topic co-occurrence.

## 4.4 Similarity in Modeling Novelty

Novelty is inversely proportional to similarity. Novel items can be thought of as ones which are in far distance not similar to the ones seen (experienced) before [3,82]. Thus by reversing the similarity calculation we can come up with a measurement for novelty. Distance in a vector space is computed by using different metrics for similarity like cosine similarity and Euclidean distance. Cosine similarity is a well-known method for identifying distance in a multidimensional space applied when the representation model is in a vector space. Towne et al. [41] use cosine similarity of LDA representations of short text documents to find the similarity (but not the novelty) of documents and present a method for validating the algorithm against human perceptions of similarity. When the representation model is in a vector space, cosine similarity is an ideal similarity metric for calculating similarity [21].

With the Topic Model representation we can have a multidimensional space, so we can measure cosine similarity and distance and then apply it in modeling different properties of a text document; in our study that property is the novelty of the research publications. Wang et al. [8] use a similar approach to assess the online idea novelty by applying the LDA topic model. We use the STM topic model and describe how the computational novelty of research publication abstracts can be modeled using the cosine similarity of Topic Modeling representation of the abstracts. The overall model for novelty based on Topic Modeling and similarity is shown in Figure 4.7.

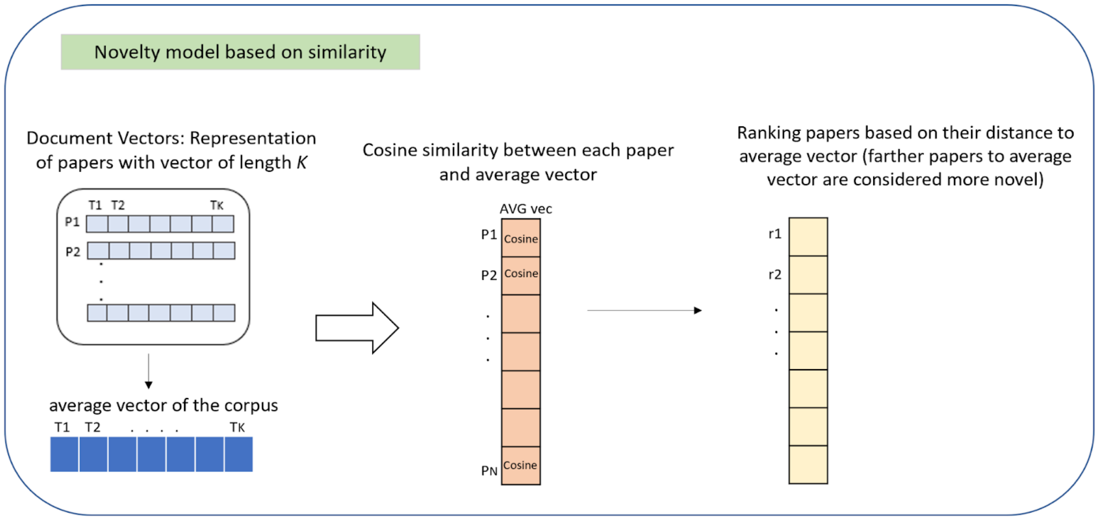



Figure 4.7: Novelty model based on similarity.

For each paper in the corpus, the Topic Model algorithm delivers the topic proportion of all topics in the corpus. Thus, as discussed in section 4.2, each paper is considered as a $K$ dimensional vector, where $K$ is the number of topics initially specified in the model (20 in our study). Each indexed element of the document's vector corresponds to the proportion of the document that is composed of the topic with the same index number. By computing the average of the vectors for all papers in the corpus, we obtain an average vector ($Avg$) with the same dimension, $K$. We assign a novelty score to each paper by measuring the distance between the vector of the paper and the average vector of the corpus. We use cosine similarity for measuring the distance to the average vector for each paper. We then rank papers based on their distance to the average vector. Papers that are farther from the average vector have lower similarity and are considered more novel. Equation 4.2 shows the formula for measuring novelty of the paper $P_i$ based on this approach.

$$Novelty\ P_i = \frac{1}{cosine\ similarity\,(P_i\,\&\,Avg)} \tag{4.2}$$

Following is the stepwise process of this computational novelty model that is based on similarity. Each paper $P_i$ is represented by a vector of $K$ dimensions, where $K$ is the number of topics discovered by Topic Model:

$$P_i : \{T_{i,1}\, , T_{i,2}\, , \ldots , T_{i,K}\, \},$$

where $T_{i,K}$ is a number between 0 and 1 representing the proportion of the paper $P_i$ that is composed of Topic $K$. We average the vectors of the papers in the corpus, where the averaging will be performed per dimension, so the average vector $Avg$ will also have $K$ dimensions:

$$Avg : \{T_{a,1}\, , T_{a,2}\, , \ldots , T_{a,K}\, \},$$

where $T_a$,$K$ is a number between 0 and 1 representing the average proportion of all papers for Topic $K$. Given two vectors of $K$ dimension for any paper $P_i$ and the average vector $Avg$, the cosine similarity between the paper $P_i$ and the average vector $Avg$ is calculated using a dot product and magnitude as shown in equation 4.3 [9]. By reversing this value for each paper we get the novelty score for the paper.

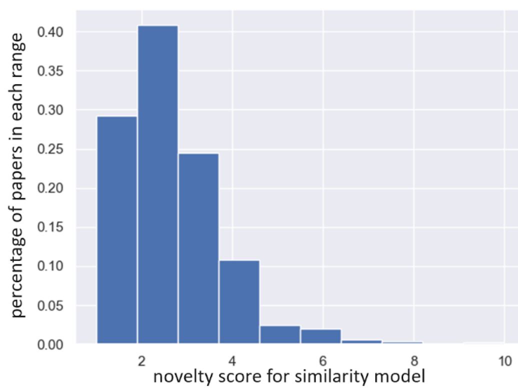
$$cosine\ similarity\ \ P_i\,\&\,Avg\ = \frac{P_i \cdot Avg}{||P_i||\ ||Avg||} = \frac{\sum_{t=1}^{K} (T_{i,t} \cdot T_{a,t}\,)}{\sqrt{\sum_{t=1}^{K} (T_{i,t}\,)^2}\ \sqrt{\sum_{t=1}^{K} (T_{a,t}\,)^2}} \tag{4.3}$$

In this section, we described novelty as being inversely proportional to similarity. In other words, we can say novelty is proportional to the distance from the corpus average, such that high novelty is associated with higher distance to the corpus average and low novelty is associated with lower distance to the corpus average. In the next section, we demonstrate the application and results of running the novelty model based on similarity on our dataset.

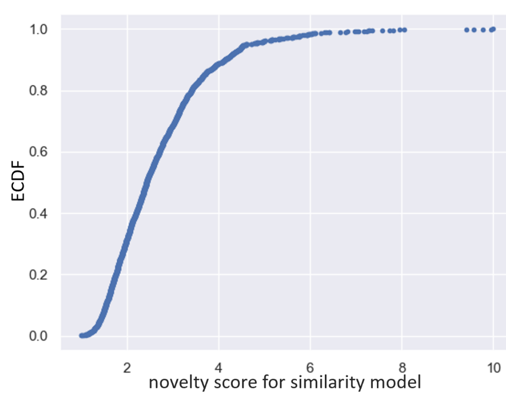
### 4.4.1 Application and Results for Novelty Using Similarity

We developed the second novelty model using the similarity approach and output from the Topic Model on our dataset of abstracts. The second novelty model measures the distance of each document's content to the average of the contents of the corpus.

Based on this model, we assigned a novelty score to each paper and ranked the papers according to their novelty scores similar to the process done for topic co-occurrence/combination approach (see section 4.3). Figure 4.8 shows the distribution of novelty scores based on the similarity approach. We can see that the novelty score for the majority of the papers falls into the first few ranges/bins. These are actually the papers closer to the corpus average. There are not many papers in the right side ranges. Those are in far distance to the corpus average and considered as more novel papers. Parts (a) and (b) of Figure 4.8 illustrate respectively the probability density function and cumulative distribution function of novelty scores.




(a) Probability density function for the similarity model




(b) Cumulative distribution function of novelty scores

Figure 4.8: Distribution of novelty scores based on the similarity approach.

Similar to the process for the first novelty model, we represent the three most novel papers, three from the moderate novel papers, and three of the least novel papers in the dataset as examples. These three sets of papers are displayed in Tables 4.5, 4.6, and 4.7 respectively. Again, for picking the three moderate novel papers, we randomly selected three papers within the middle 40% of the sorted list of papers based on their novelty score.

Table 4.5: The three most novel paper abstracts in the corpus.

| **Title** | **Novelty score** | **Most prominent Topic** |
|---|---|---|
| 1: "Put Your Warning Where Your Link Is: Improving and Evaluating Email Phishing Warnings" (2019) | 10 | 13 (web security, user interfaces) |
| 2: "What Do We Really Know about How Habituation to Warnings Occurs Over Time?: A Longitudinal fMRI Study of Habituation and Polymorphic Warnings" (2017) | 9.94 | 13 (web security, user interfaces) |
| 3: "To Miss is Human: Information-Theoretic Rationale for Target Misses in Fitts' Law" (2017) | 9.76 | 13 (web security, user interfaces) |

By investigating the three most novel papers picked by the similarity novelty model [83–85], shown in table 4.5, we found that the most prominent Topic for all of these papers is identical to the least prominent Topic of the corpus average, which is Topic 13, associated with web security and user interfaces. Moreover, the topic proportion distributions for these three papers are very different from the corpus average. That is because in this model the algorithm picks the least similar (farthest) papers to the corpus average as the most novel ones.

The first novel paper in table 4.5 suggests effective methods to improve email phishing warning designs to reduce phishing click through rates compared to current approaches [83]. The second paper presents two methods of eye tracking and fMRI (functional magnetic resonance imaging) for measuring habituation, "*decreased response to a repeated warning*", to investigate how habituation develops over time [84]. The paper suggests eye tracking as a valid measure of the mental process of habituation to warnings [84]. It also suggests polymorphic warning design as an effective method for preventing habituation compared to conventional warning [84]. The third paper is a more theoretical paper with many mathematical formulas describing theory and methods in HCI research and user interfaces from a new perspective [85].

Table 4.6: Three of the moderate novel paper abstracts in the corpus.

| **Title** | **Novelty score** | **Most prominent Topic** |
|---|---|---|
| 1: "Beyond the Patient Portal: Supporting Needs of Hospitalized Patients" (2019) | 3.06 | 6 (health) |
| 2: "Let's Play Together: Adaptation Guidelines of Board Games for Players with Visual Impairment" (2019) | 2.87 | 3 (visualization, accessibility) |
| 3: "In the Eye of the Student: An Intangible Cultural Heritage Experience, with a Human-Computer Interaction Twist" (2018) | 2.43 | 2 (smart home/life, smart city) |

The first moderate novel paper shown in Table 4.6 presents design suggestions for patient portal technologies to support the needs of hospitalized patients that existing patient portals do not support [86]. The most prominent topic in this paper is Topic 6, which is related to health. The ideas presented in this paper help future patient portals to engage hospitalized patients and caregivers in their health care [86].

The second paper focuses on accessibility (Topic 3) for visually impaired players. It describes a user-centered design approach to providing accessibility and equal chances of victory for visually impaired players by developing adaptation guidelines to make board games irrespective of the user's visual impairment [87]. The third paper in Table 4.6 discusses postcolonialism and decolonization of HCI and technology designing in non-western communities and environments, and designing applications to document intangible heritage by engaging students from non-western cultural communities [88]. Topic Model identified Topic 2, which is associated with smart home/life, smart city, and communities, as the most prominent topic for this paper. The topic proportion vectors of all of these three papers are neither very far nor very close to the corpus average (all are in kind of average distance).

Table 4.7: The three least novel paper abstracts in the corpus.

| **Title** | **Novelty score** | **Most prominent Topic** |
|---|---|---|
| 1: "Evaluating the Effect of Feedback from Different Computer Vision Processing Stages: A Comparative Lab Study" (2019) | 1 | 4 (Interaction design, ethic) |
| 2: "Experiential Augmentation: Uncovering The Meaning of Qualitative Visualizations when Applied to Augmented Objects" (2018) | 1.02 | 4 (Interaction design, ethic) |
| 3: "Towards Collaboration Translucence: Giving Meaning to Multimodal Group Data" (2019) | 1.04 | 4 (Interaction design, ethic) |

The first paper in Table 4.7, picked by the similarity algorithm as the least novel paper in the corpus, reports the challenges of designing interactions for pattern matching algorithms and the role of visual feedback to help the users to understand the related application's operation correctly [89]. The second paper suggests enhancing the interaction design language to help digital interfaces to better communicate in a real world context by considering qualitative visualizations [90]. The third paper discusses team members' multimodal interaction with each other and artifacts, both in online and face-to-face team work, and describes approaches to making visible the features of group activities [91]. The first prominent topic of all of the three least novel papers is identical to the most prominent topic of the corpus average, that is Topic 4 (interaction design, ethics). The topic proportion distribution for these papers are close to the corpus average.

For all the 3 papers in Table 4.5, (most novel papers) the most prominent (or top) topic in these papers is identical to the least prominent topic in the corpus average, and overall the topic vectors of these papers were in a far distance of the corpus average in the vector space. These papers, which are considered as points in the vector space, are indeed the outliers picked by the computational novelty algorithm as the most novel papers in the corpus. On the other hand, we see that the most prominent topic for all three least novel papers listed in Table 4.7 is identical to the most prominent topic of the corpus average and their topic vectors are very close to the corpus average. We also observed that papers closer to the corpus average have more evenly distributed topic proportions, compared to the farther distance papers which have one or a few topic(s) with high topic proportion and their remaining topics have very low topic proportion. We did not find any meaningful correlation between the novelty score and the year of the publication for the similarity based novelty approach.

Sections 4.3 and 4.4 addressed research question 3 concerning how computational novelty models of research publications can/may be defined by using Topic Modeling, based on the topic combination and similarity approaches. In the next section, we compare and analyze the results from the two novelty models.

## 4.5 Comparative Analysis of the Two Computational Models of Novelty in HCI Papers

By comparing the results obtained from the two novelty models, we observe that the two models have different novelty score distributions. The table shown in Figure 4.9 displays the summary statistics of novelty scores in two approaches and Figure 4.10 shows the cumulative distribution functions of these two approaches in one plot. The ECDFs in Figure 4.10 expose clear differences among the novelty score distributions of the two novelty models. To make sure that the observed difference is significant and is not due to the random chance (or just for this specific dataset), we performed a hypothesis significance testing on these results and found that the novelty score distribution of the two novelty models are statistically significantly different with a very small $p$ value less than 0.0001 (i.e. the two approaches have statistically significantly different novelty score distributions). This suggests that there are fundamental differences in the novelty measurement of the different approaches for modeling novelty while all (each model) can be true/correct but each one is looking at the concept of novelty from a different aspect/perspective (which can/may lead to the different consequences in terms of creativity). This suggests that the novelty model has a lot to do with the novelty score. We can conclude that not all/different novelty models point to the same thing. The novelty model based on topic combination considers only the most prominent/representative parts of each item in measuring the novelty of an item while the similarity based novelty model looks at the item as a whole and compares its similarity with the average of the dataset/corpus.

| | Topic combination | Similarity | difference between two approaches |
|---|---|---|---|
| mean | 5.5 | 2.7 | 2.8 |
| median | 5.3 | 2.4 | 2.9 |
| standard deviation | 1.3 | 1.1 | 0.2 |

Figure 4.9: Summary statistics of novelty scores in two novelty approaches.

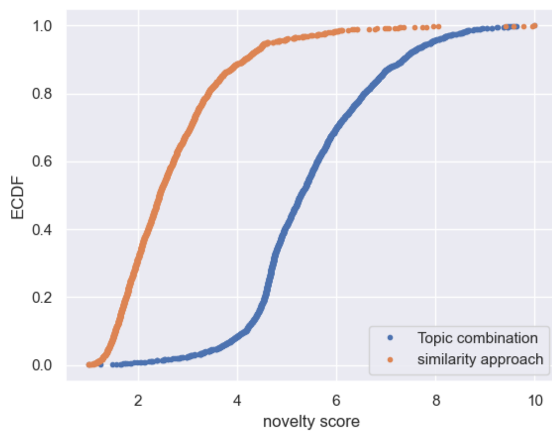



Figure 4.10: Comparing the ECDFs of novelty scores for the two approaches.

We see that the mean is larger in the topic combination/co-occurrence model, and the variance does appear larger as well. According to the table shown in Figure 4.9 and Figure 4.10, the similarity based novelty model has less absolute variability in novelty scores than the topic combination/co-occurrence model. By looking at Figures 4.5a and 4.8a (see sections 4.3 and 4.4 respectively), we see that the novelty scores for the topic combination approach are more distributed in different distinct ranges compared to the similarity approach in which we see that most novelty scores are accumulated in the first four ranges at left and the probability density function is somewhat left skewed. These are the papers closer to the centroid and we have a few highly novel papers (outliers) in this model. The similarity approach seems to be more appropriate for when we just need to pick a few numbers of most novel papers in the whole corpus. On the other hand, the topic combination approach is more appropriate for when we want to pick both the most novel and least novel papers. The distribution of the topic combination novelty model is close to a normal distribution. In the similarity approach, we have a large number of less novel papers accumulated in the first few ranges on the left side of the PDF plot (Figure 4.8 a). Novelty scores are too close to each other in these ranges that labeling a paper as least novel may not be very reasonable. However, in the topic combination approach, both most novel and least novel papers are discernible because we have more distinct novelty scores in this model. This suggests that the topic combination approach gives us the desired novelty score distribution (is more appropriate) for when we want to rank the papers based on the novelty score (i.e. when we want to have distinct ranking), and for when we want to identify both most novel and least novel papers. But the similarity approach may not be a good choice for ranking because the novelty scores of papers in the first couple of ranges are very close to each other.

From another aspect, we can say that when we are interested in the most prominent parts of an item for the novelty (i.e. when we are looking at the novelty from the aspect of novel combination of most prominent parts/topics for each item/paper), topic combination approach may be more apt to be applied. But in the cases that we want to look at an item as a whole and we want to consider all (even not prominent or trivial) parts of the item, the similarity based approach seems to be a better choice. Indeed, the similarity based approach is used when we are looking at the novelty from the perspective of how unsimilar/far is the content of the paper as a whole to the average of the rest papers in the corpus. So it mostly depends on what is our (application) goal and what is our perspective of novelty and from which aspects we are looking at the concept of novelty. We should see what novelty means to us and to our application, and also what parts of the item are important for us to be considered in terms of novelty computation (most prominent/representative parts or the whole item?). Do we want to have a distinct ranking with clear most novel and least novel papers, or do we just want to pick the most novel papers in a corpus? Each of these may direct us to a different novelty model.

Section 4.5 addressed research question 4 concerning how topic combination and topic similarity measure novelty differently on the same corpus of research papers. How the novelty score distribution differs, and how the meaning of novelty may be expressed differently in the two models: co-occurrence (combination) and similarity of topics.

## 4.6 A Study Comparing Computational Models of Novelty with Human Perception of Novelty

We designed and conducted a study to answer the fifth research question in this dissertation concerning how computational models of novelty compare to human perception of novelty.

This user study was designed to help understand how each of the topic co-occurrence and similarity based computational models of novelty compare with human perception of novelty. The degree of novelty in our computational novelty models was determined in two ways: topic combination (co-occurrence), by the degree of atypicality (novelty) of topic combinations in paper abstract, and similarity, by the degree of distance between the paper abstract vector and the corpus average vector. We recruited 9 faculty, senior PhD students, and alumni with a background in HCI as the experts in the domain of our dataset (HCI) to complete our study and give their feedback. As discussed by Blandford et al. in [92], analyzing data from 4 to 12 participants could be sufficient for qualitative studies like this. While 9 participants does not allow a statistical analysis of significance, it is sufficient for a qualitative analysis of the human perception of novelty in the different conditions of the study.

### 4.6.1 Methods and Participants

This study included 9 participants recruited from the College of Computing and Informatics at a large comprehensive public university in North America, including 4 faculty, 4 senior PhD students, and 1 PhD alumnus. Gender distribution was 3 males and 6 females. All the participants had a background in HCI as the experts in the domain of our dataset (HCI).

The study includes 3 sections where each participant completes 3 surveys, one survey in each section. The first section focuses on the first novelty model, topic combination. The second section is similar to the first section but focuses on the second novelty model which is based on similarity. The third section considers both novelty models.

In this study, we used an online google survey that enables users to complete the study at their convenient time and location. The instructions for the task for each section are displayed at the beginning of the survey. Participants are first introduced to the study and its goal. The tasks of the entire survey, along with the recommended break time for each section, are described to the participants. In each section, we asked the participants to read a few paper abstracts selected by the novelty algorithm from our dataset of HCI papers, and then answer some questions. The estimated time for completing the overall study is between 30 to 60 minutes.

At the beginning of the survey, we first briefly introduced our research and the purpose of our study as follows:
"Thank you for participating in this survey. The goal of this study is to explore how computational models of novelty match human perception of novelty. This study shows abstracts of research publications in the ACM Digital Library in the field of Human Centered-Computing (HCC) that have been rated for novelty based on a computational model, and your task is to provide the human perception of novelty for these publications."

In survey one, we first ask the participants to read a high novel paper abstract picked by our first novelty model (topic combination approach) and ask them to answer the following questions. The purpose of this part of the survey is to find whether the paper picked as novel by computational novelty algorithm seems novel by human experts as well. We also want to inspect the reason why human experts select a paper abstract as novel. Participants were not aware of the novelty rating assigned to the papers by the novelty algorithm. The first part of survey one is as follows:

1.1 Please read the following research publication abstract and answer the following questions.

> "***Keppi: A Tangible User Interface for Self-Reporting Pain***
> *Motivated by the need to support those managing chronic pain, we report on the iterative design, development, and evaluation of Keppi, a novel pressure-based tangible user interface (TUI) for the self-report of pain intensity. In-lab studies with 28 participants found individuals were able to use Keppi to reliably report low, medium, and high pain as well as map squeeze pressure to pain level. Based on insights from these evaluations, we ultimately created a wearable version of Keppi with multiple form factors, including a necklace, bracelet, and keychain. Interviews indicated high receptivity to the wearable design, which satisfied additional user-identified needs (e.g., discreet and convenient) and highlighted key directions for the continued refinement of tangible devices for pain assessment.*"

1.1.1 Do you find this paper novel? If yes, why?

1.1.2 How novel is this paper with respect to topic combinations? Do you find it novel because of new combination of topics you see in it?

1.1.3 Please highlight the words, phrases, or sentences giving you the sense of novelty (if you find any) in the abstract of the paper while reading it. Then list the highlighted words here.

1.1.4 What is your main area of expertise in the domain of HCI?

1.1.5 Is the presented paper in your area of expertise?

After the first part of survey one, we presented one high novel and one low novel paper abstract and asked the participants to read them without letting them know the novelty rate of the papers. We then asked them to select the more novel paper among the two based on their perception of novelty, and answer the followed questions. The second part of survey one is as follows (here, paper (b) is the high novel paper, and paper (a) is the low novel one based on the topic co-occurrence novelty algorithm):

1.2 Please read the following two research publication abstracts and answer the followed questions.

Paper a)

> "***Technology and the Givens of Existence: Toward an Existential Inquiry Framework in HCI Research***
> *The profound impact of digital technologies on human life makes it imperative for HCI research to deal with the most fundamental aspects of human existence. Arguably, insights from existential philosophy and psychology are highly relevant for addressing such issues. Building on previous attempts to bring in existential themes and terminology to HCI, this paper argues that Yalom's notion of "the givens of existence", as well as related work in experimental existential psychology, can inform the development of an existential inquiry framework in HCI. The envisioned framework is intended to complement current approaches in HCI by specifically focusing on the existential aspects of the design and use of technology. The paper reflects on possible ways, in which existential concepts can support HCI research, and maintains that adopting an existential framework in HCI*

*would be consistent with the overall conceptual development of the field.*"

Paper b)

"***SmartManikin: Virtual Humans with Agency for Design Tools***
*When designing comfort and usability in products, designers need to evaluate aspects ranging from anthropometrics to use scenarios. Therefore, virtual and poseable mannequins are employed as a reference in early-stage tools and for evaluation in the later stages. However, tools to intuitively interact with virtual humans are lacking. In this paper, we introduce SmartManikin, a mannequin with agency that responds to high-level commands and to real-time design changes. We first captured human poses with respect to desk configurations, identified key features of the pose and trained regression functions to estimate the optimal features at a given desk setup. The SmartManikin's pose is generated by the predicted features as well as by using forward and inverse kinematics. We present our design, implementation, and an evaluation with expert designers. The results revealed that SmartManikin enhances the design experience by providing feedback concerning comfort and health in real time.*"

1.2.1 Which of these two papers you find more novel? Why you find that more novel?
1.2.2 How novel is your selected paper with respect to topic combinations? Do you find it novel because of new combination of topics you see in it?
1.2.3 Please highlight the words, phrases, or sentences giving you the sense of novelty (if you find any) in the abstract of your selected novel paper while reading it. Then list the highlighted words here.
1.2.4 Are the presented papers in your area of expertise in the domain of HCI?

At the end of survey one, we recommended participants take at least 10 minutes break to refresh their minds and then proceed to the second survey. Survey 2 included the same process and questions but this time the presented papers were selected by the second novelty model (similarity approach). Similar to survey 1, in survey 2, the participants are not aware of the novelty rating assigned to the papers by the novelty algorithm. For the reader of this dissertation, in the second part of survey 2, paper (a) is the high novel paper, and paper (b) is the low novel one picked by the second novelty algorithm. Following we show survey 2 with the same parts:

2.1 Please read the following research publication abstract and answer the following questions.

> "***Put Your Warning Where Your Link Is: Improving and Evaluating Email Phishing Warnings***
> *Phishing emails often disguise a link's actual URL. Thus, common anti-phishing advice is to check a link's URL before clicking, but email clients do not support this well. Automated phishing detection enables email clients to warn users that an email is suspicious, but current warnings are often not specific. We evaluated the effects on phishing susceptibility of (1) moving phishing warnings close to the suspicious link in the email, (2) displaying the warning on hover interactions with the link, and (3) forcing attention to the warning by deactivating the original link, forcing users to click the URL in the warning. We assessed the effectiveness of such link-focused phishing warning designs in a between-subjects online experiment (n=701). We found that link-focused phishing warnings reduced phishing click-through rates compared to email banner warnings; forced attention warnings were most effective. We discuss the implications of our findings for phishing warning design.*"

2.1.1 Do you find this paper novel? If yes, why?

2.1.2 How novel is this with respect to topic combinations? Do you find it novel because of new combination of topics you see in it?

2.1.3 Please highlight the words, phrases, or sentences giving you the sense of novelty (if you find any) in the abstract of the paper while reading it. Then list the highlighted words here.

2.1.4 Is the presented paper in your area of expertise?

2.2 Please read the following two research publication abstracts and answer the following questions.

Paper a)

> "***What Do We Really Know about How Habituation to Warnings Occurs Over Time?: A Longitudinal fMRI Study of Habituation and Polymorphic Warnings***
>
> *A major inhibitor of the effectiveness of security warnings is habituation: decreased response to a repeated warning. Although habituation develops over time, previous studies have examined habituation and possible solutions to its effects only within a single experimental session, providing an incomplete view of the problem. To address this gap, we conducted a longitudinal experiment that examines how habituation develops over the course of a five-day workweek and how polymorphic warnings decrease habituation. We measured habituation using two complementary methods simultaneously: functional magnetic resonance imaging (fMRI) and eye tracking.*
>
> *Our results show a dramatic drop in attention throughout the workweek despite partial recovery between workdays. We also found that the polymorphic warning design was substantially more resistant to habituation compared to conventional warnings, and it sustained this advantage throughout the five-day experiment. Our findings add credibility to prior studies by showing that the pattern of habituation holds across a workweek, and indicate that cross-sectional habituation studies are valid proxies for longitu-*

*dinal studies. Our findings also show that eye tracking is a valid measure of the mental process of habituation to warnings.*"

Paper b)

"***Evaluating the Effect of Feedback from Different Computer Vision Processing Stages: A Comparative Lab Study***
*Computer vision and pattern recognition are increasingly being employed by smartphone and tablet applications targeted at lay-users. An open design challenge is to make such systems intelligible without requiring users to become technical experts. This paper reports a lab study examining the role of visual feedback. Our findings indicate that the stage of processing from which feedback is derived plays an important role in users' ability to develop coherent and correct understandings of a system's operation. Participants in our study showed a tendency to misunderstand the meaning being conveyed by the feedback, relating it to processing outcomes and higher level concepts, when in reality the feedback represented low level features. Drawing on the experimental results and the qualitative data collected, we discuss the challenges of designing interactions around pattern matching algorithms.*"

2.2.1 Which of these two papers you find more novel? Why you find that more novel?
2.2.2 Please highlight the words, phrases, or sentences giving you the sense of novelty (if you find any) in the abstract of your selected novel paper while reading it. Then list the highlighted words here.
2.2.3 Are the presented papers in your area of expertise in the domain of HCI?

Similar to the first survey, at the end of survey two, we recommend participants take at least 10 minutes break to refresh their minds and then proceed to the next (third) survey. In the third survey, we present one of the most novel paper abstracts picked by the topic co-occurrence algorithm (paper a) and one of the most novel paper abstracts picked by the similarity algorithm (paper b) both being in the same area (interactive games). The participants were not aware of the type of algorithm and novelty rating of these two papers. The purpose of the third survey is to find which of the two novelty algorithms may be closer to human exerts perception of novelty. The third (and last) survey is as follows:

3.1 Please read the following two research publication abstracts on the topic of interactive games. Then answer the following question.

Paper a)

> "***Supporting Easy Physical-to-Virtual Creation of Mobile VR Maze Games: A New Genre***
>
> *With the fast development of virtual reality games, one of the key research questions is how players may express their creativity and participate in the process of game design. In this paper, we present a new game genre which combines user-controlled game design in physical space with game play in virtual space on a mobile device. The new system supports authoring by anyone, creating virtual reality games that can be easily modified or developed for physical space, and be used anywhere by novice end-users without any knowledge of tracking technology. We present the design and implementation of the system, as well as a user experiment. Findings illustrate that the proposed system promotes participation and provides a richer, more interactive and engaging experience.*"

Paper b)

> "***An Odd Kind of Pleasure" : Differentiating Emotional Challenge in Digital Games***
>
> *Recent work introduced the notion of emotional challenge as a means to*

*afford more unique and diverse gaming experiences. However, players' experience of emotional challenge has received little empirical attention. It remains unclear whether players enjoy it and what exactly constitutes the challenge thereof. We surveyed 171 players about a challenging or an emotionally challenging experience, and analyzed their responses with regards to what made the experience challenging, their emotional response, and the relation to core player experience constructs. We found that emotional challenge manifested itself in different ways, by confronting players with difficult themes or decisions, as well as having them deal with intense emotions. In contrast to more'conventional' challenge, emotional challenge evoked a wider range of negative emotions and was appreciated significantly more by players. Our findings showcase the appeal of uncomfortable gaming experiences, and extend current conceptualizations of challenge in games.*"

3.1.1 Which of the two paper abstracts above you find (more) novel? Please explain why?

### 4.6.2 Qualitative Analysis

We analyzed the participants' responses to the survey questions to understand how computational models of novelty compare to human perception of novelty for the two novelty models.

Our data showed that 7/9 participants found the novel paper picked by the first novelty algorithm (topic combination), as a novel paper (survey question 1.1.1). Regarding the second novelty model (similarity approach), 5/9 participants found the novel paper picked by the novelty algorithm as a novel paper (question 2.1.1). When asking the participants whether they found the paper abstract novel with respect to topic combinations, the majority of the participants (6/9) answered "yes" for the first model (question 1.1.2), while regarding the second novelty model (similarity), the majority (7/9) answered "no" (question 2.1.2). This is promising as the basis of the novelty calculation for the first model is combination of topics while that is not the case regarding the second novelty model. It shows our first novelty model reasonably picked the paper abstract with novel topic combinations.

Regarding comparing a high novel and a low novel paper picked by the two novelty algorithms in the second part of surveys 1 and 2 (questions 1.2.1 and 2.2.1), 5/9 participants explicitly mentioned in survey 1 that they give a higher rate of novelty to paper b, that is the paper picked as high novel by the first novelty algorithm. 2/9 participants found paper a more novel than paper b. P4 commented: "*Paper a is more novel to me because I have a better understanding of what the aim is. Paper b is much less clear to me. It's hard to evaluate something as novel if you don't really understand it. I don't know understand what 'agency' the mannequin has been given, so the second one is very unclear.*". 2/7 participants expressed they can not judge which paper is more novel between the two, or gave both the same novelty rating. For instance, P7 mentioned: "*From the abstracts, I find them to have a similar degree of novelty based on prior experience at the time of the survey without additional literature review. I will select paper (a - existential inquiry framework) as having a slight novelty edge. I believe that this is because (a) I am less familiar with the existential philosophy components vs the kinematic and analytics components, and (b) development of sound theoretical constructs is less frequently addressed and often presents greater challenges. It is a close call, however, and the judgment is more about feeling of novelty, as the first reflects on potential without validation through use or user study, whereas the second notes evaluative feedback*". Pertaining to the second novelty algorithm (similarity based), 4/9 participants found paper a (the one rated as high novel by the algorithm) more novel than paper b (the one rated as low novel by the algorithm). 2/9 participants (P1) selected paper b as the more novel one. P6 found none of the papers novel stating: "*None of these papers I find novel, as*

*I have encountered papers like these earlier*". Again 2/9 participants expressed they can not judge which paper is more novel between the two, or gave both the same novelty rating.

5/9 of the participants stated the papers presented in survey 1 by the first novelty algorithm are in their area of expertise in the domain of HCI (questions 1.1.5 and 1.2.4). Regarding the second novelty model in survey 2, 4/9 participants stated the presented papers are in their area of expertise in the domain of HCI (questions 2.1.4 and 2.2.3).

Regarding the third survey (question 3.1.1), 5/9 participants selected paper a (the one rated as high novel by the topic combination algorithm) as the more novel one compared to paper b (the one rated as high novel by the similarity algorithm). P4 gave both papers a and b the same novelty rating with more tendency to paper a. P4 expressed: "*I find both of these to be quite novel. The first one is novel in developing a new system that allows users to 'build as they go' in a VR game. I also find this one quite interesting and relevant to a current ongoing research project. However, the first abstract is sparse on details, it's very high level, so that makes it a bit hard to evaluate. The second abstract also seems somewhat novel, though I believe there has already been a fair amount of work in 'designing uncomfortable experiences', though it is not my research area. I have definitely seen lots of work on designing for discomfort in terms of digital interactive theater experiences to help users develop empathy. If I had to pick which one is 'most novel' I would probably pick the first*". 2/9 participants gave a higher novelty rating to paper b picked by the similarity algorithm. P9 had a notable comment: "*It's really hard to compare the novelty of these two papers since they are novel in different ways. The first one presents a novel system that allows players to design games, which is novel since it breaks the traditional bounds of who gets to make games. The second one is just a survey, which is not a novel research technique, but the topic of the survey is rather novel since it tries to capture something about* *the emotional aspect of games. Overall, I'm more excited by the idea of emotional challenges in games, so I'll say Paper B is more novel*". This participant found both papers rated as high novel by the two novelty algorithms as novel papers. 1/9 participants (P5) had difficulty judging between the 2 papers and did not have a selection.

#### 4.6.2.1 Thematic Analysis

To understand the main reasons of the participants (human experts) for selecting a paper as novel and interesting, we analyzed the participants' responses to the survey questions. We performed a thematic analysis of the responses the participants gave to the survey questions. Overall, four main themes were found from the survey answers.

- Novel approach, tool, application, or design
- Demonstrating good experiment, evaluation, new findings and study results
- Usefulness
- Personal experience and relevance

In this section, we elaborate on each of these themes.

**Novel approach, tool, application, or design**

Most participants stated one of the reasons they selected a paper abstract as novel was because of finding a novel approach, tool, application, or design in it. P6 found the paper in section 1.1.1 of the survey as novel stating, "*I never knew any tangible User Interface for reporting Pain*". This participant found the application and user interface introduced in this paper abstract novel mentioning: "*novel pressure-based tangible user interface*". P7 commented: "*It presents an application - wearable, tangible interface for reporting pain level that is different from other kinds of wearable tangible interfaces that I have encountered...*". This participant shows how a new application and interface that is different from the previous ones they saw give them the sense of novelty of this paper. P7 continued: "*The combination for the specific application - pressure sensor being validated as a pain reporting instrument and user experience design for pain tracking - I find to be novel*". P4 stated: "*because I don't know of other work that approaches this problem using tangibles*". This participant found the paper novel because of the novel approach used in it for a problem.

P1 found the second paper (paper b) in section 1.2.1 of the survey as the novel one and explained: "*the authors in this paper introduced a new (novel) technology with a unique name (SmartManikin) with the design, implementation, and evaluation. This is something I did not find in the first paper (Technology and the Givens of Existence: Toward an Existential Inquiry Framework in HCI Research). It seems like in the first paper the authors only reviewed the literature and reflected on this. In addition, the first paper seemed theoretical and the second paper was practical*". This participant mentioned the right point. The second paper, which was rated by the novelty algorithm as a highly novel one as well, introduces a novel design and technology while the first paper rated as a low novel one by the algorithm, just reviewed and reflected on the existing literature. P6 also found paper b in section 1.2.1 as the novel one because of novel development. This participant expressed: "*paper b as it's a new development rather than adding to an existing one*". P2 found the first paper (paper a) in section 2.2.1 as the more novel paper compared to the second paper (paper b) because of the interesting and novel methods applied in it. This participant commented: "*the paper used two interesting methods to measure habituation, fMRI and eye tracking*" as one of their reasons for finding this paper novel.

P1 selected paper a (the paper picked by the topic combination novelty model as highly novel) as the more novel one compared to paper b (the paper picked by the similarity novelty model as highly novel) in section 3.1.1 because of presenting a novel design and introducing a new concept. This participant expressed: "*In paper a the* *authors clearly presented a new thing (new genre) for game design. They explicitly mention this is a new genre. There are also certain words that made me think this is more novel (present, design, implementation, proposed, promotes, richer, engaging). This is not something I found in paper b. Paper b did not introduce a new concept, but rather used what was introduced in other work on emotional challenge*". Similarly, P4 found the first paper (paper a) in section 3.1.1 more novel than the second one (paper b) because of developing a novel system. This participant described "*the first one is novel in developing a new system that allows users to 'build as they go' in a VR game ...*" which shows one of their reasons to find a paper as novel is seeing a novel system is developed.

**Demonstrating good experiment, evaluation, new findings and study results**

Another theme we found in the several participants' responses to survey questions for why they found a paper novel was good evaluation and study results. P3 found the paper presented in section 1.1.1 of the survey as novel because "*it shows good results when users interacted with a wearable version of Keppi to measure their pain level*". This participant also highlighted the phrases "*In-lab study*" and "*pain assessment*" among the ones giving them the sense of novelty in the abstract of the paper while reading it. Another participant, P2, found this paper novel because it was mentioned in the paper abstract that "*interviews indicated high receptivity*". P7 also found this paper novel commenting: "*The following sentences provide an indication of evaluative steps in the combination development, which is the foundation for the view of novelty: - In-lab studies with 28 participants found individuals were able to use Keppi to reliably report low, medium, and high pain as well as map squeeze pressure to pain level. - Interviews indicated high receptivity to the wearable design, which satisfied additional user-identified needs (e.g., discreet and convenient) and highlighted key directions for the continued refinement of tangible devices for pain assessment.*". This participant indicates that one of the foundations for the view of novelty is evaluative steps in combination development. This shows that demonstrating good evaluation and study results is important in human (experts) perception of the novelty of a paper. P3 found the second paper (paper b) more novel than the first one (paper a) in section 1.2.1 of the survey and commented: "*because it involves a user study that evaluates the interaction between users and SmartManikin*". P9 explicitly pointed to the new evaluation method in the first paper (paper a) of section 2.2.1 as one of the reasons they found this paper as novel. This participant commented: "*the authors in Paper A developed a new evaluation method (eye tracking), which also makes the paper novel*".

P1 found the paper presented in section 2.1.1 of the survey as novel and expressed: "*Yes. Although warning messages are not a new thing, but I have not seen previous research that evaluates the effectiveness of the location of the warning message and how it appears. I am not aware of any research that did this before*". As stated by this participant, the reason they found this paper novel is what was evaluated in the paper, not the main concept of it. Similarly, P2 found this paper novel and mentioned "*between-subjects online experiment (n=701)*" among the phrases giving them the sense of novelty in the abstract of the paper while reading it. P2 expressed one of the reasons they found paper a in section 2.2.1 more novel than paper b as "*I found the first paper more novel as it assessed an intervention, polymorphic warning design, with a longitudinal study that previous papers did not do. Longitudinal studies are more appropriate when checking user behavior with everyday applications*". This participant also mentioned: "*The finds are more interesting than the other paper. This paper also supports the previous work in the area with more strong methodology and findings*". These examples show that participants considered the evaluation methods and study results as one of their criteria for selecting a paper as novel.

**Usefulness**

Another theme we found among the responses of participants is usefulness. P2 commented in section 1.1.1 of the survey as: "*... If Keppi is reliable to identify pain intensity, it would be a really awesome tool for the patients as well as for the doctors. Additionally tangible and wearable interface is easy to use and convenient for a large portion of the population. I believe a lot of people will benefit from this research*". P2 also expressed usefulness as one of the reasons they found the second paper (paper b) in section 1.2.1 as the more novel one. This participant expressed: "*I find this paper novel as SmartManikin can provide human-like feedback to design changes in real-time concerning comfort and health. It can be a useful tool in creative tasks for feedback. This research can also be useful in human-AI co-creativity and creativity support tool research*". P3 found the first paper (paper a) in section 3.1.1 of survey more novel than the second paper (paper b) and commented: "*because it shows how users may express their creativity during the game design...*". This participant added: "*They provided design implementations which improve the game design*". Similarly, P4 found the first paper more novel than the second paper. This participant mentioned the useful system in the paper as one of his reasons for finding this paper as novel. They expressed: "*The first one is novel in developing a new system that allows users to 'build as they go' in a VR game ...*". P2 also found the first paper more novel and commented: "*I found the first abstract more novel. Both abstracts are about digital/virtual gaming but the first abstract seems more timely and useful for players, especially for novices. The first research is unique as the users can design the game in physical space by anyone and does not require any knowledge. It also shows findings about promoting participation and engaging experience. This research can lead to future directions in VR games*". These examples demonstrate that usefulness is among the reasons human experts may find a paper novel.

**Personal experience and relevance**

Personal experience is another theme we found in the participants' answers to the survey questions. Some of the participants selected a paper as novel because they could relate the subject of the paper to their personal research or life. P2 expressed one of the reasons they found the paper in section 1.1.1 of the survey as a novel paper is his personal experience. This participant commented: "*I found this paper novel... I personally have a hard time describing my pain intensity to my doctor each and every time as I think it is subjective ...*". They found this research beneficial for him. P7 found the paper in section 2.1.1 somewhat but not very novel. This participant commented: "*I find this paper to be a qualified, somewhat novel... In my experience this seems more akin to intelligent user interface patterns for trust or persuasive recommendation, and I have seen quite a few*". P4 pointed to his personal experience as well in selecting paper a as the more novel paper in section 3.1.1. This participant described "*The first one is novel... I also find this one quite interesting and relevant to a current ongoing research project... The second abstract also seems somewhat novel, though I believe there has already been a fair amount of work in 'designing uncomfortable experiences', though it is not my research area. I have definitely seen lots of work on designing for discomfort in terms of digital interactive theater experiences to help users develop empathy. If I had to pick which one is 'most novel' I would probably pick the first*". These examples show how personal experience and relevance may affect a human expert participant's decision in evaluating the novelty of a paper.

The survey results showed the first novelty model (topic combination) is more reasonably close to human perception of novelty compared to the second novelty model (similarity). However, it is still far from claiming a significant match or alignment with human perception of novelty considering limitations in the study including the number of participants, and human limitations in reading thousands of papers in a corpus. From one perspective, it can be said that computational models of novelty can/may be able to calculate the novelty of research publications in a corpus better than human because of human limitations in reading large number of papers in a corpus. The themes we found by performing thematic analysis on the participants' responses to the survey questions, showed that the criteria for evaluating the novelty of a research publication may slightly differ in humans compared to the computational algorithms. That is because of limitations and different backgrounds and experiences related to humans. However, this does not mean declining either one. Each of these two has its own advantages and disadvantages. Computational models of novelty and human experts in the field can be complement to each other in many applications including peer review process. Computational models presented in this dissertation are not capturing things that are personal to the person. They are only capturing topics and their relevance in the corpus, not connecting it to the person. That is because the focus in this dissertation is on objective novelty, that is computational novelty of documents in a corpus, not personalized/subjective novelty. Considering the themes found from qualitative analysis, a human expert may consider different and more aspects of novelty which is lacking in computational models. Computational models have the advantage of better representation of the whole corpus but they do not capture what the human relevance, expertise, and personal preferences are. And that is important because future work could be to have computational models of novelty better responsive to individual and personal understandings.

### 4.6.3 Limitations

One limitation of the study comparing computational models of novelty with human perception of novelty, is the number of participants. The participants had to be selected from a specific category, that is faculty or PhD students with a concentration in the domain of HCI (i.e. experts in HCI), and we did not have many options to select and invite as the participants for our study. Because of the small number of participants, the results may not generalized to represent the whole population, and it is a limitation. Another limitation is a sampling of the population. Because of the small number of participants, the demographics of the participants may not be representative of the demographics of experts in HCI.

# CHAPTER 5: SUMMARY AND FUTURE WORKS

Novelty, as a major component of creativity, does not have a unique definition. There can be different definitions of novelty, and consequently different models of novelty, depending on different perspectives/aspects being considered. In order to choose the appropriate novelty model for an application, it is important to consider the aspect or perspective by which we define the novelty in the domain of the application.

In this dissertation, we first defined a framework through which different novelty models can be reviewed, categorized, and compared. Then we described the Pique system to demonstrate the usefulness of our framework and computational models of novelty in educational recommender systems. We demonstrated how systems like Pique can help students to expand their knowledge. We then discussed how using Topic Modeling for representing data may facilitate computational novelty modeling of research publications. In the Topic Modeling approach for representing data, the distribution of topics in each document gives a unified vector representation for each document with a much smaller vector dimension compared to the other approaches for representing data such as "bag of words" and TF-IDF vector representations. In those approaches, we have a high dimensional and sparse document representation while in Topic Modeling the vectors are low dimension that are highly interpretable, providing a structured representation for text understanding. It would not be very efficient to do the novelty measurements if we had a large dimensional vector space like in the bag of words and TF-IDF. Topic Modeling reduces the dimension of the vector of variables in large dimensional spaces and gives an interpretable structure. Topic Modeling is a meaningful and effective dimension reduction process and gives us the potential of exploring more novelty measurements for future research. This approach to representing text documents enables us to consider multiple documents and produce a structure automatically in a representation that can be used in novelty measurement. These properties of Topic Modeling facilitate modeling novelty more effectively considering time and space complexities.

Representing research publications with Topic Modeling appears to be consistent and comprehensible enough for applying different novelty measurements. Research publications are different from news, recipes, or other kinds of textual data, as peer reviewed research publications are accepted on the basis of a novel contribution. Scientific papers are typically in-depth, despite the other sources of textual data. Topic Modeling automates the process of generating the representation (and discovering hidden topics in a large corpus) for measuring novelty, whereas other approaches such as applying keywords in research papers ( [9, 26]) and ingredients in recipes ( [13]) rely on humans to generate the representation. In approaches applying author-defined keywords, the authors have no idea what is in the other papers, but in the topic modeling approach, the topics are based on what is in the other papers in the corpus. This makes Topic Modeling as an appropriate choice for representing research publications for measuring their novelty.

We developed and compared two different novelty models for measuring computational novelty in research publications using Topic Modeling for papers in a dataset of about 2,000 HCI publications. One model measures novelty with respect to the most prominent/representative components (topics) of the item (research publications) and the atypicality of each combination of components in that item/document. The second model measures novelty with respect to all components constituting the item and its dissimilarity (distance) to the corpus average. We applied data analysis and statistical inference to learn about novelty scores produced by the two different novelty models. We did an exploratory analysis and discussed the most and least novel papers (as well as some of the middle ones) picked by each of the two novelty models. Each presented novelty model has its own pros and cons, each capturing one aspect of novelty and capable of identifying some surprising-seeming papers that the other missed.

The similarity based model measures how far the vector of all components of a publication are from the corpus average, and the topic combination model measures how unusual the combination of the most prominent/representative parts of the publication are compared to the rest of the dataset. The similarity model may not be a good choice when we want to rank the whole dataset and need distinct ranking. It can be a good choice when we want to pick the few most novel papers in the whole dataset as it actually identifies the outliers. The topic combination model is a better choice for when we want to rank all the publications in the dataset having distinct ranking. Deciding which model of novelty we should apply depends on what our application is and also which aspect of novelty is of more importance to us (our application). We found that the novelty scores distributions for the two novelty models are statistically significantly different, which suggests that different novelty models may not be pointing to the same things as novel and that not all novelty models give us the same novelty score. Different novelty models may point to the different things or different items as novel because they are looking at the novelty from different perspectives. In order to pick the right novelty model appropriate for our application, it is very important to be clear about what aspect of novelty is of importance.

We designed and conducted a study comparing computational models of novelty with human perception of novelty. We recruited 9 faculty, senior PhD students, and PhD alumni as the experts in the domain of our dataset, HCI, and asked them to provide their feedback in an online survey after reading some paper abstracts selected by computational models of novelty. The qualitative analysis of the results suggests that the first novelty model (topic combination) is more reasonably closer to human perception of novelty compared to the second novelty model (similarity). However, we do not claim that there is a significant match or alignment with human perception of novelty, considering limitations in the study including the number of participants, and human limitations in reading thousands of papers in a corpus. From one aspect, we can say that computational models of novelty can/may be able to calculate the novelty of research publications better than humans because of human limitations in reading large numbers of papers in a corpus.

We found four main themes by performing thematic analysis on the participants' responses to the study as: 1)novel approach, tool, application, or design, 2)good experiment, evaluation, and study results, 3)usefulness, and 4)personal experience and relevance. The themes we found showed that the criteria for evaluating novelty of a research publication in humans may not be a complete match with the computational algorithms. That is because of limitations and different backgrounds and experiences related to humans. However, we believe these two, computational models of novelty and human experts in the field, can complement each other in many applications including the peer review process. Computational models and human perception of novelty both have advantages and disadvantages. Computational models of novelty have the advantage of seeing the whole corpus at a glance which is a limitation for human to read thousands of papers in a corpus. But the advantage of human perception is it may consider different and more aspects of novelty which is lacking in computational models. Thus we think these two can be good complements to each other in the peer review process and that will be one of our focuses in future research works. Another focus for our future work would be developing personalized computational models of novelty (p-novelty) considering human expertise and experience in the model.

Human perception of novelty (even experts) is limited when having big repositories/corpus of thousands text documents. It is impossible for one person to read several thousands of text documents in a corpus to evaluate the novelty of each document or discover the most novel text documents among all. Computational models of novelty are developed to overcome this challenge. Computational models of novelty can scan a whole corpus of thousands of documents and assign a novelty score to each document. We do not have any ground truth available for modeling the novelty of research publications. This research is a first step to a better understanding of the role of computational novelty in describing what we mean by novelty in a specific application and comparing that to other models of novelty. This is enabled by our framework and by our definition of two models of novelty.

In our future research, we plan to study new and more comprehensive novelty models by combining different novelty models each of which measures a different aspect of novelty. Novelty evaluation is one of the criteria in the peer review process and as discussed human experts are not able to read and recall all thousands of papers in a dataset for a particular domain of study to be able to evaluate if a new/single paper is novel compared to the rest of the corpus. In our future work, we are interested in examining the possibility of using effective computational novelty models as an assistant and complement (but not as substitution) to the human experts in the peer review process for evaluating the novelty of scientific papers.

# REFERENCES


[1] D. M. Blei, "Probabilistic topic models," *Communications of the ACM*, vol. 55, no. 4, pp. 77–84, 2012.

[2] K. Grace, M. Maher, D. Fisher, and K. Brady, "A data-intensive approach to predicting creative designs based on novelty, value and surprise," *International Journal of Design, Creativity, and Innovation*, 2014.

[3] M. L. Maher, "Evaluating creativity in humans, computers, and collectively intelligent systems," in *Proceedings of the 1st DESIRE Network Conference on Creativity and Innovation in Design*, pp. 22–28, 2010.

[4] M. L. Maher and D. H. Fisher, "Using ai to evaluate creative designs," in *DS 73-1 Proceedings of the 2nd International Conference on Design Creativity Volume 1*, 2012.

[5] P. Mishra and D. Henriksen, "A new approach to defining and measuring creativity: Rethinking technology & creativity in the 21st century," *TechTrends*, vol. 57, no. 5, p. 10, 2013.

[6] T. M. Amabile, "Social psychology of creativity: A consensual assessment technique.," *Journal of personality and social psychology*, vol. 43, no. 5, p. 997, 1982.

[7] K. Grace and M. L. Maher, "What to expect when you're expecting: The role of unexpectedness in computationally evaluating creativity.," in *ICCC*, pp. 120–128, Ljubljana, 2014.

[8] K. Wang, B. Dong, and J. Ma, "Towards computational assessment of idea novelty," in *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 2019.

[9] M. Mohseni, M. L. Maher, K. Grace, N. Najjar, F. Abbas, and O. Eltayeby, "Pique: Recommending a personalized sequence of research papers to engage student curiosity," in *International Conference on Artificial Intelligence in Education*, pp. 201–205, Springer, 2019.

[10] L. Macedo and A. Cardoso, "Towards artificial forms of surprise and curiosity," in *Proceedings of the European Conference on Cognitive Science, S. Bagnara, Ed*, pp. 139–144, 1999.

[11] K. Grace, M. L. Maher, D. Wilson, and N. Najjar, "Personalised specific curiosity for computational design systems," in *Design Computing and Cognition'16*, pp. 593–610, Springer, 2017.

[12] X. Niu, "Curious george: Modeling surprise to inspire curiosity." 2019.

[13] K. Grace, M. L. Maher, N. Davis, and O. Eltayeby, "Surprise walks: Encouraging users towards novel concepts with sequential suggestions," in *Proceedings of the 9th International Conference on Computational Creativity (ICCC 2018). Association of Computational Creativity*, 2018.

[14] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1532–1543, 2014.

[15] P. Baldi and L. Itti, "Of bits and wows: A bayesian theory of surprise with applications to attention," *Neural Networks*, vol. 23, no. 5, pp. 649–666, 2010.

[16] X. Niu, F. Abbas, M. L. Maher, and K. Grace, "Surprise me if you can: Serendipity in health information," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2018.

[17] M. Mohseni and M. L. Maher, "A framework for exploring computational models of novelty in unstructured text," in *Proceedings of the 6th International Conference on Information System and Data Mining*, ICISDM '22, (New York, NY, USA), p. 36â45, Association for Computing Machinery, 2022.

[18] S. Siddiqui, M. L. Maher, N. Najjar, M. Mohseni, and K. Grace, "Personalized curiosity engine (pique): A curiosity inspiring cognitive system for student directed learning.," in *CSEDU (1)*, pp. 17–28, 2022.

[19] P. Adamopoulos and A. Tuzhilin, "On unexpectedness in recommender systems: Or how to better expect the unexpected," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 5, no. 4, pp. 1–32, 2014.

[20] L. D. Baker, T. Hofmann, A. McCallum, and Y. Yang, "A hierarchical probabilistic model for novelty detection in text," in *Proceedings of International Conference on Machine Learning*, vol. 254, Citeseer, 1999.

[21] A. Verheij, A. Kleijn, F. Frasincar, and F. Hogenboom, "A comparison study for novelty control mechanisms applied to web news stories," in *2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, vol. 1, pp. 431–436, IEEE, 2012.

[22] J. Allan, R. Gupta, and V. Khandelwal, "Topic models for summarizing novelty," in *ARDA Workshop on Language Modeling and Information Retrieval. Pittsburgh, Pennsylvania*, Citeseer, 2001.

[23] L. R. Varshney, F. Pinel, K. R. Varshney, A. Schörgendorfer, and Y.-M. Chee, "Cognition as a part of computational creativity," in *2013 IEEE 12th international conference on cognitive informatics and cognitive computing*, pp. 36–43, IEEE, 2013.

[24] R. G. Morris, S. H. Burton, P. M. Bodily, and D. Ventura, “Soup over bean of pure joy: Culinary ruminations of an artificial chef.,” in *ICCC*, pp. 119–125, Citeseer, 2012.

[25] B. Uzzi, S. Mukherjee, M. Stringer, and B. Jones, “Atypical combinations and scientific impact,” *Science*, vol. 342, no. 6157, pp. 468–472, 2013.

[26] N. Carayol, L. Agenor, and L. Oscar, “The right job and the job right: Novelty, impact and journal stratification in science,” *Impact and Journal Stratification in Science (March 5, 2019)*, 2019.

[27] V. Shwartz and I. Dagan, “Still a pain in the neck: Evaluating text representations on lexical composition,” *Transactions of the Association for Computational Linguistics*, vol. 7, pp. 403–419, 2019.

[28] G. Salton and M. J. McGill, *Introduction to modern information retrieval.* New York, NY: Mcgraw-Hill, 1983.

[29] Z. S. Harris, “Distributional structure,” *Word*, vol. 10, no. 2-3, pp. 146–162, 1954.

[30] J. Brownlee, “How to use word embedding layers for deep learning with keras,” *On October 4, 2017 in Deep Learning for Natural Language Processing*, 2017.

[31] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent dirichlet allocation,” *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022, 2003.

[32] D. Sarkar, “Traditional methods for text data..” https://towardsdatascience.com/understanding-feature-engineering-part-3-traditional-methods-for-text-data-f6f7d70acd41, 2018.

[33] E. Gabrilovich, S. Dumais, and E. Horvitz, “Newsjunkie: providing personalized newsfeeds via analysis of information novelty,” in *Proceedings of the 13th international conference on World Wide Web*, pp. 482–490, 2004.

[34] M. Karkali, F. Rousseau, A. Ntoulas, and M. Vazirgiannis, “Efficient online novelty detection in news streams,” in *Web Information Systems Engineering–WISE 2013: 14th International Conference, Nanjing, China, October 13-15, 2013, Proceedings, Part I 14*, pp. 57–71, Springer, 2013.

[35] J. Firth, “A synopsis of linguistic theory, 1930-1955,” *Studies in linguistic analysis*, pp. 10–32, 1957.

[36] J. Brownlee, “What are word embeddings for text?.” https://machinelearningmastery.com/what-are-word-embeddings/, 2019.

[37] T. Mikolov, W.-t. Yih, and G. Zweig, “Linguistic regularities in continuous space word representations,” in *Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics: Human language technologies*, pp. 746–751, 2013.

[38] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

[39] Wikipedia, "Word2vec." https://en.wikipedia.org/wiki/Word2veccite$_{r}ef$ $-$ *mikolov*.

[40] T. Kekeç, L. van der Maaten, and D. M. Tax, "Pawe: Polysemy aware word embeddings," in *Proceedings of the 2nd International Conference on Information System and Data Mining*, pp. 7–13, 2018.

[41] W. B. Towne, C. P. Rosé, and J. D. Herbsleb, "Measuring similarity similarly: Lda and human perception," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 8, no. 1, pp. 1–28, 2016.

[42] M. E. Roberts, B. M. Stewart, D. Tingley, E. M. Airoldi, *et al.*, "The structural topic model and applied social science," in *Advances in neural information processing systems workshop on topic models: computation, application, and evaluation*, vol. 4, pp. 1–20, Harrahs and Harveys, Lake Tahoe, 2013.

[43] T. M. Cover and J. A. Thomas, "Information theory and statistics," *Elements of information theory*, vol. 1, no. 1, pp. 279–335, 1991.

[44] C. C. Aggarwal and C. C. Aggarwal, "Proximity-based outlier detection," *Outlier analysis*, pp. 111–147, 2017.

[45] M. A. Pimentel, D. A. Clifton, L. Clifton, and L. Tarassenko, "A review of novelty detection," *Signal processing*, vol. 99, pp. 215–249, 2014.

[46] C. E. Shannon, "A mathematical theory of communication," *The Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.

[47] G. Jumarie, "Relative informationâwhat for?," in *Relative Information*, pp. 1–11, Berlin, Heidelberg: Springer, 1990.

[48] K. Grace, M. L. Maher, M. Mohseni, and R. PEREZ Y PEREZ, "Encouraging p-creative behaviour with computational curiosity," ICCC, 2017.

[49] A. Barto, M. Mirolli, and G. Baldassarre, "Novelty or surprise? front psych. 2013; 4: 907," 2013.

[50] D. Kokotsaki, V. Menzies, and A. Wiggins, "Project-based learning: A review of the literature," *Improving schools*, vol. 19, no. 3, pp. 267–277, 2016.

[51] I. Guk and D. Kellogg, "The zpd and whole class teaching: Teacher-led and student-led interactional mediation of tasks," *Language Teaching Research*, vol. 11, no. 3, pp. 281–299, 2007.

[52] D. M. Blei and J. D. Lafferty, "A correlated topic model of science," *The annals of applied statistics*, vol. 1, no. 1, pp. 17–35, 2007.

[53] G. Bouma, "Normalized (pointwise) mutual information in collocation extraction," *Proceedings of GSCL*, vol. 30, pp. 31–40, 2009.

[54] L. Vygotsky, "Interaction between learning and development," *Readings on the development of children*, vol. 23, no. 3, pp. 34–41, 1978.

[55] D. E. Berlyne, "Curiosity and exploration," *Science*, vol. 153, no. 3731, pp. 25–33, 1966.

[56] D. A. Schön, *The reflective practitioner: How professionals think in action*. England, UK: Routledge, 2017.

[57] D. A. Kolb, *Learning style inventory*. Boston, MA: McBer and Company, 1999.

[58] J. Cowan, *On Becoming an Innovative University Teacher: Reflection in Action: Reflection in Action*. United Kingdom: Society for Research Into Higher Education & Open University Press, 2006.

[59] A. Jones and K. Issroff, "Learning technologies: Affective and social issues in computer-supported collaborative learning," *Computers & Education*, vol. 44, no. 4, pp. 395–408, 2005.

[60] J. R. Hill and S. M. Land, "Open-ended learning environments: A theoretical framework and model for design.," 1998.

[61] Wikipedia, "Topic model." https://en.wikipedia.org/wiki/Topic$_{m}odel,$ 2022.

[62] L. Hulstaert, "Topic model." https://www.datacamp.com/tutorial/lda2vec-topic-model, 2017.

[63] M. Aznag, M. Quafafou, and Z. Jarir, "Correlated topic model for web services ranking," *International Journal of Advanced Computer Science and Applications*, vol. 4, no. 6, 2013.

[64] J. Aitchison, "The statistical analysis of compositional data," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 44, no. 2, pp. 139–160, 1982.

[65] M. E. Roberts, B. M. Stewart, and D. Tingley, "Stm: An r package for structural topic models," *Journal of Statistical Software*, vol. 91, no. 1, pp. 1–40, 2019.

[66] Y. Zhang, J. Callan, and T. Minka, "Novelty and redundancy detection in adaptive filtering," in *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 81–88, 2002.

[67] S. Kaplan and K. Vakili, "The double-edged sword of recombination in breakthrough innovation," *Strategic Management Journal*, vol. 36, no. 10, pp. 1435–1457, 2015.

[68] M. E. Roberts, B. M. Stewart, D. Tingley, C. Lucas, J. Leder-Luis, S. K. Gadarian, B. Albertson, and D. G. Rand, "Structural topic models for open-ended survey responses," *American journal of political science*, vol. 58, no. 4, pp. 1064–1082, 2014.

[69] C. Lucas, R. A. Nielsen, M. E. Roberts, B. M. Stewart, A. Storer, and D. Tingley, "Computer-assisted text analysis for comparative politics," *Political Analysis*, vol. 23, no. 2, pp. 254–277, 2015.

[70] J. Chang, "Package 'lda'." http://cran.r-project.org/web/packages/lda/lda.pdf, 2015.

[71] M. Taddy, "Multinomial inverse regression for text analysis," *Journal of the American Statistical Association*, vol. 108, no. 503, pp. 755–770, 2013.

[72] K. Grace, M. L. Maher, D. Fisher, and K. Brady, "Modeling expectation for evaluating surprise in design creativity," in *Design Computing and Cognition'14*, pp. 189–206, Springer, 2015.

[73] A. T. Adams, E. L. Murnane, P. Adams, M. Elfenbein, P. F. Chang, S. Sannon, G. Gay, and T. Choudhury, "Keppi: A tangible user interface for self-reporting pain," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2018.

[74] C.-Y. Hsu, R. Hristov, G.-H. Lee, M. Zhao, and D. Katabi, "Enabling identification and behavioral sensing in homes using radio reflections," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2019.

[75] B. Lee, T. Jin, S.-H. Lee, and D. Saakes, "Smartmanikin: virtual humans with agency for design tools," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2019.

[76] M. Eslami, S. R. Krishna Kumaran, C. Sandvig, and K. Karahalios, "Communicating algorithmic process in online behavioral advertising," in *Proceedings of the 2018 CHI conference on human factors in computing systems*, pp. 1–13, 2018.

[77] J. van Kollenburg, S. Bogers, E. Deckers, J. Frens, and C. Hummels, "How design-inclusive uxr influenced the integration of project activities: three design cases from industry," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 1408–1418, 2017.

[78] C. Qu, C. Sas, and G. Doherty, "Exploring and designing for memory impairments in depression," in *Proceedings of the 2019 CHI conference on human factors in computing systems*, pp. 1–15, 2019.

[79] M. Asad, C. A. Le Dantec, B. Nielsen, and K. Diedrick, "Creating a sociotechnical api: Designing city-scale community engagement," in *Proceedings of the 2017 CHI conference on human factors in computing systems*, pp. 2295–2306, 2017.

[80] V. Kaptelinin, "Technology and the givens of existence: Toward an existential inquiry framework in hci research," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2018.

[81] S. Hauser, R. Wakkary, W. Odom, P.-P. Verbeek, A. Desjardins, H. Lin, M. Dalton, M. Schilling, and G. De Boer, "Deployments of the table-non-table: A reflection on the relation between theory and things in the practice of design research," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2018.

[82] R. Saunders and J. S. Gero, "Designing for interest and novelty: Motivating design agents," in *Computer Aided Architectural Design Futures 2001: Proceedings of the Ninth International Conference held at the Eindhoven University of Technology, Eindhoven, The Netherlands, on July 8–11, 2001*, pp. 725–738, Springer, 2001.

[83] J. Petelka, Y. Zou, and F. Schaub, "Put your warning where your link is: Improving and evaluating email phishing warnings," in *Proceedings of the 2019 CHI conference on human factors in computing systems*, pp. 1–15, 2019.

[84] A. Vance, B. Kirwan, D. Bjornn, J. Jenkins, and B. B. Anderson, "What do we really know about how habituation to warnings occurs over time? a longitudinal fmri study of habituation and polymorphic warnings," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 2215–2227, 2017.

[85] J. Gori, O. Rioul, and Y. Guiard, "To miss is human: Information-theoretic rationale for target misses in fitts' law," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 260–264, 2017.

[86] S. Haldar, S. R. Mishra, M. Khelifi, A. H. Pollack, and W. Pratt, "Beyond the patient portal: supporting needs of hospitalized patients," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2019.

[87] F. da Rocha Tomé Filho, P. Mirza-Babaei, B. Kapralos, and G. Moreira Mendonça Junior, "Let's play together: Adaptation guidelines of board games for players with visual impairment," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–15, 2019.

[88] D. Giglitto, S. Lazem, and A. Preston, "In the eye of the student: An intangible cultural heritage experience, with a human-computer interaction twist," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2018.

[89] J. Kittley-Davies, A. Alqaraawi, R. Yang, E. Costanza, A. Rogers, and S. Stein, "Evaluating the effect of feedback from different computer vision processing stages: a comparative lab study," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2019.

[90] D. Lo, D. Lockton, and S. Rohrbach, "Experiential augmentation: Uncovering the meaning of qualitative visualizations when applied to augmented objects," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–8, 2018.

[91] V. Echeverria, R. Martinez-Maldonado, and S. Buckingham Shum, "Towards collaboration translucence: Giving meaning to multimodal group data," in *Proceedings of the 2019 chi conference on human factors in computing systems*, pp. 1–16, 2019.

[92] A. Blandford, D. Furniss, and S. Makri, *Sampling and Recruitment*, pp. 23–31. Cham: Springer International Publishing, 2016.