What We Talk About When We Talk About Games: Bottom-Up Game Studies Using Natural Language Processing

James Owen Ryan[1,2], Eric Kaltman[1], Michael Mateas[1], and Noah Wardrip-Fruin[1]
[1] Expressive Intelligence Studio
[2] Natural Language and Dialogue Systems Lab
University of California, Santa Cruz
{jor, ekaltman, michaelm, nwf}@soe.ucsc.edu

ABSTRACT

In this paper, we endorse and advance an emerging bottom-up approach to game studies that utilizes techniques from natural language processing. Our contribution is threefold: we present the first complete review of the growing body of work through which this methodology has been innovated; we present a latent semantic analysis model that constitutes the first application of this fundamental bottom-up technique to the domain of digital games; and finally, unlike earlier projects that have only written about their models, we introduce and evaluate a tool that serves as an interface to ours. This tool is GameNet, in which nearly 12,000 games are linked to the games to which they are most related. From an expert evaluation, we demonstrate that, beyond being an interface to our model, GameNet may be used more generally as a research tool for game scholars. Specifically, we find that it is especially useful for the scholar who wishes to explore a relatively unfamiliar area of games, but that it may also be used to discover unforeseen cases related to topics that have already been thoroughly researched.

Categories and Subject Descriptors

I.5.3 [**Clustering**]: Similarity Measures; K.8.0 [**General**]: Games

General Terms

Algorithms, Design, Measurement, Human Factors

Keywords

game studies, literature review, natural language processing, research tools, machine learning, latent semantic analysis

1. INTRODUCTION

There is a nearly boundless accumulation of text about digital games that exists online in structured collections ripe for analysis. What insights could we glean, from these trillions of words, about videogames as a medium? If we were to harness the sheer volume of all this language, what could we build? We believe that the application of natural language processing (NLP) techniques to text about games represents a hugely promising but relatively unexplored area of game-studies research. This approach is bottom-up in the sense that it yields findings that emerge (often unexpectedly) from extensive language use, whereas in the more conventional top-down approach a scholar starts from a preconceived notion that she then attempts to substantiate. Bottom-up methods are particularly useful in exploratory research, and for many topics they also represent a more grounded approach. In game studies, there are numerous subjects of interest, including games themselves, that cannot be logically organized into neatly discrete categories. As a result, top-down approaches to these subjects tend to organize their research and discussions around concepts like game genre, which is a problematic borrowing from game-industry marketing practice. These contrived notions restrict not only innovation of the form, but also our discussion of it.

We might also consider how other disciplines have had great insights using similar bottom-up methods. In the digital humanities, there has been in the last decade an outpouring of work applying NLP techniques to various text collections, to great results. As an example, the authors of [52] processed over 3,000 library-science dissertations, dating from 1930 onward, to generate an overview of the field that shed new light on its origins, its history, and its outlook. We wonder what we could learn about our (albeit much younger) field from a study applying this very method to a collection of game-studies dissertations. In [35], 80,000 articles and advertisements from a colonial U.S. newspaper were processed in an effort that afforded better understanding of early American society. This method applied to a corpus of early videogame magazines could be equally illuminating. The toolset for this methodology is well-developed and there is certainly no shortage of text about games—the outstanding matter is doing the work.

While this area of game-studies research is vastly unexplored, there are a handful of scholars that have begun to venture into it. As part of our contribution here, we provide in this paper the first complete survey of this growing literature. But while this foundational work has been successful in producing findings that top-down methods could not have, there is a recurrent issue that has inhibited the emerging methodology. None of the models that have been developed so far can be engaged beyond the prose of their respective publications, and this is highly problematic. Due to the inherent complexity of such models, it is difficult to adequately describe them, or even to give a sense of their general implications, with prose alone. It is our belief that bottom-up models produced by NLP techniques must be visualized or made interactive in some way.

From this impetus, we present (and evaluate) GameNet, the first such interactive visualization of a game-studies NLP model. As a tool in which nearly 12,000 games are linked to one another according to how related they each are, GameNet could not have been built using top-down methods. Underpinning it is a model that was developed by processing a collection of Wikipedia articles about games, totaling some 14.5M words, using an NLP technique called latent semantic analysis (LSA). In addition to our literature review and GameNet, the third facet to our contribution here is the first application of LSA, one of the foundational bottom-up NLP techniques, to the domain of digital games. Above all, we hope that this project will spur new and interesting research using the bottom-up approach to game studies that we describe and practice herein.

In the following section, we provide a review of prior game-studies work that has used NLP, as well as a detailed description of latent semantic analysis (both of which assume an audience that is new to NLP). In Section 3, we recount the extraction and preparation of our collection of thousands of Wikipedia articles describing individual games, as well as our derivation of a latent semantic analysis model trained on that collection. GameNet and its expert evaluation are discussed in Section 4. Finally, we conclude in Section 5.

2. BACKGROUND

There is a growing body of work in which NLP techniques are employed in game-studies research, centered in large part around the efforts of José Zagal, Noriko Tomuro, and their (former) colleagues at DePaul University. More precisely, this work is characterized by its application of techniques from *statistical natural language processing*, a subfield of NLP in which bottom-up statistical methods are applied to large collections of natural-language text. In this section, we provide the first complete review of this literature, before explaining latent semantic analysis, the statistical NLP technique driving our current work. Throughout, we attempt to explain these concepts in such a way that readers who are not NLP practitioners may understand them.

2.1 Statistical NLP for Game Studies

In [56], the first project to use statistical NLP for game studies, Zagal and Tomuro study the specific language used to evaluate games across a collection of nearly 400,000 game reviews submitted by users to the website GameSpot [1]. First, they gather 723 unique adjectives that modify the word 'gameplay' in some review, and then, treating these adjectives as the core vocabulary with which game appraisal is expressed, proceed to examine them more deeply according to the contexts they occur in. Specifically, they compile the 5000 words that most frequently appear either directly before or directly after the adjectives. From here, they represent each adjective by its distribution with respect to these various contexts—using machine learning parlance, this is the *feature representation* that they use—and proceed to *cluster* the adjectives. Clustering is a procedure whereby objects are grouped together such that ones in the same cluster are more similar to one another (with regard to their feature representations) than to objects in other clusters. For this task, the authors use k-*means* [27], one of the standard clustering algorithms. Here, k is a *hyperparameter*—a parameter whose value is set by the user prior to runtime (as opposed to a parameter whose value is 'learned' by the algorithm itself during runtime)—that specifies how many clusters the algorithm will partition the input set of objects into. After some initial exploration, the authors set k to 30, and then use a subset of these 30 adjective clusters to propose a typology of what they call "primary elements of gameplay aesthetics" (12). Using this, they attest the existence of a rich language for appraising aesthetic aspects of gameplay, but note that the specific vocabulary used by players appears to be different from that employed by scholars and designers.

We note that this particular finding would not have been reached using a top-down approach. The primary elements that they present are rooted in specific language attested across thousands of game reviews, not in preconceived notions that the authors set out to test. In fact, they emphasize their surprise that the concept of 'emergence' did not appear as a primary element. This highlights a fundamental appeal of bottom-up scholarship, which is that it often yields unexpected insights.

[58] is a journal article in which Zagal, Tomuro, and Shepitsen argue for the use of NLP in game-studies research using three example studies. Here, we outline only the first of these, as we present the others elsewhere. In this brief study, the authors apply *readability metrics* to 1500 professionally written game reviews extracted from GameSpot. A readability metric is a formula used to determine, ostensibly, the level of education needed to understand a text. Typically, these formulae operate on the number and length of the syllables, words, and sentences of a text. Using three common metrics—*SMOG* [30], the *Coleman-Liau index* [15], and the *Gunning fog index* [6]—the authors find that the reviews are written at a secondary-education reading level. From these results, they argue against criticism that game reviews are written poorly and for a young demographic.

In [43], Raison and others extract fine-grained player appraisals of games (found in amateur reviews) and use these to cluster the games themselves. These fine-grained appraisals are in the form of *co-clusters* derived from the listing in [56] of 723 adjectives (that modified 'gameplay' in a review) and the contexts they occurred in, which we described above. Whereas in standard clustering one set of objects (all of the same type) is partitioned into clusters of similar objects, in co-clustering two sets (having different types of objects) are simultaneously partitioned such that the elements of a cluster in the first set are bonded by being similarly associated with the elements of a particular cluster in the other set. So, what is produced is a set of co-clusters, rather than a set of regular clusters. In the case of the study at hand, each of the authors' 3000 derived co-clusters comprises a cluster of adjectives and a cluster of contexts such that those particular adjectives all tend to occur in those particular contexts and, likewise, those contexts all tend to feature those adjectives. As an example, one of the co-clusters they list has {'great', 'amazing', 'excellent', ...} as its adjectival cluster and {'graphics', 'look', 'sound',

...} for its contextual cluster. Extrapolating from these co-clusters, as well as statistical associations between clusters of the same type, the authors argue about player perceptions of games more generally. For instance, they suggest that games that are perceived as being addictive, fun, or exciting are also perceived as being unique, deep, and innovative. Finally, the authors use their co-clusters as a feature representation with which to represent games themselves, which they then cluster using k-means. That is, they represent a game by a *feature vector* that specifies how many times particular adjectives were used to evaluate particular gameplay aspects in reviews for that game. (For more on using feature vectors to represent things in the world, see Section 2.2.) From their clustering analysis, they observe (among other things) that clusters could not always be understood at the level of gameplay—for example, they cite a cluster of games that came from different gameplay genres but that were each based on animated television series.

As before, we find that the very nature of these results is rooted in the authors' bottom-up method of inquiry. The fact that some of their clusters included games from multiple conventional genres highlights a key argument for this approach—when games are clustered according to how people actually talk about them, the resulting bottom-up typology contradicts the dominant top-down one.

The task of extracting fine-grained player opinions about games, found in the previous study, can be characterized as belonging to a subfield of NLP called *sentiment analysis* (SA). The general aim of SA is to automatically extract subjective information, such as opinions, from texts. Another endeavor in this area is [13], in which Chiu and others process a corpus of over 200,000 Chinese-language reviews of mobile games to investigate how *sentiment polarity* (what percentage of a text is positive in sentiment) differs across review portions pertaining to different categories of game appraisal. Their core contribution is a novel opinion-extraction procedure that is tailored to handle Chinese-specific nuances that challenge techniques developed using English-language text, but their analysis reveals some interesting findings as well. For instance, of the five appraisal categories that the text of each of their reviews pertains to—*gameplay*, *aesthetics*, *musicality*, *stability*, and *developer*—they find that the latter two are far more likely to command negative sentiment than are the first three. Briefly, we will mention that there has been other, earlier work at the intersection of SA and digital games (*e.g.*, [17], [9]), but in the interest of space, and especially because these are not game-studies contributions, we do not discuss them in detail here.

As part of a larger exploration of cultural differences in game appraisal, in [57] Zagal and Tomuro study lexical differences between Western and Japanese game reviews. Specifically, for 221 games released in both the US and Japan, they compare the nouns most frequently occurring in user reviews submitted to GameSpot to those most frequently used in user reviews submitted to GameWorld, a Japanese website. Among other differences, they observe that Japanese reviews are more critical of technical issues, while replayability appears to be more central to Western concerns.

In direct follow-up work to [43], Meidl, Lytinen, and Raison use the former's co-cluster feature representation to build a game *recommender system* [31]. A recommender system is software that predicts what else a user may like given what they are already known to like [46]. So while [43] used co-cluster feature vectors to represent games so that they could be clustered, the authors here use these same feature vectors to represent the games a person likes for the purpose of automatically recommending other games they may like. Knowing three games that a particular player likes, the system recommends the games whose feature vectors are most similar to those of the liked games. This concept is fairly complex, so we refrain from discussing it more deeply until the next section. To measure the system's accuracy, the authors employ an offline method that is conventionally used to evaluate recommender systems. From this, they report 0.86 precision—that is, 86% of the games their system recommended were indeed liked by the players being recommended to. Independently of [31], two other game recommender systems were introduced in 2014 [50, 11]. As these do not represent game-studies contributions and are unrelated to the research outlined above, we invite the interested reader to see [47], a related project in which we give an overview of this work and present an LSA-fueled recommender system of our own (which we use to test the intuitive notion that people tend to like related games).

In [19], Grace conducts two lexical analyses of developer descriptions of mobile games. After compiling and analyzing the 38 distinct verbs used in developer descriptions of 70 best-selling games across the five most popular genres in Apple's App Store, he offers three higher-level game-verb categories: verbs of *elimination* ('shoot', 'kill', 'destroy', ...), *categorization* ('match', 'separate', 'choose', ...), and *transformation* ('move', 'jump', 'rotate', ...). In the second study, Grace compares the language used in Amazon descriptions of the 20 best-selling adult-fiction books of 2011 and 2012 to Apple App Store descriptions for that platform's 20 best-selling games for those years. From these admittedly small samples, his findings suggest that books may include more violent, morbid content than games do.

Finally, in a series of recent papers published in the human-computer interaction (HCI) community [59, 60, 61, 62, 63, 64], Zhu and Fang (and others) process game reviews using a lexical approach similar to that of [56] (though they appear unaware of this earlier work). These authors conceive of their method as a refinement of earlier lexical approaches that in psychology led to the formulation of the famous five-factor model of personality [29]. From a collection of 696,801 game reviews submitted by users to GameSpot, IGN [3], and GameStop.com [2], they compile the 4,843 most frequently occurring adjectives. Using the popular lexical database WordNet [33], they merge together all synonymous adjectives to yield 788 adjective groups. Next, they proceed to represent each adjective group by a feature vector specifying which documents adjectives from that group occurred in. From here, they submit these adjective-group vectors to a statistical technique called *factor analysis* [20]. In factor analysis, statistical patterns among a set of observed variables (in this case, the adjectives) are exploited to construct a much smaller set of unobserved variables, called *factors*, that can still explain the full data set quite well. The idea is that the factors will represent core, higher-level concepts that underpin the data domain; as such, some form of factor analysis is often used in exploratory research that works bottom-up from a large amount of data. (Indeed, latent semantic analysis is itself driven by a variant of factor analysis that we introduce in Section 3.2.) The authors here find six factors (each with its own set of associated adjectives from the full data set)—which they hand label as *playability*, *creativity*, *usability*, *competition*, *sensation*, and *strategy*—and argue that, as these factors are attested in the extensive language use of game players, they could greatly inform game-design practice. In extensions to this work, they propose classifying games using these factors [59]; construct new factors using both adjectives and nouns [62]; and use adjectives associated with their *playability* factor to support playability heuristics that had been proposed by earlier HCI games work, and to submit new ones suggested by the factor [60].

Our current project is situated among the work outlined above. Like these projects, the model underpinning ours could only be built using NLP and machine-learning techniques—it would not be feasible to hand-code (using a top-down approach) representations for several thousand games. That being said, we present a novel innovation of the methodology represented by the above projects. While the majority have processed game reviews—a text domain that is inherently evaluative in its tone and purpose—we use encyclopedic text, which is more objectively descriptive in tone and more ontological in purpose. Though beyond the scope of this work, there are interesting comparisons to be made between models built using the same technique, but by processing text from different domains. As a major advantage of the particular text source we use, our model includes several thousand more games spanning a larger historical period. Furthermore, the particular technique that we use is novel. As we have mentioned above, this is the first application of latent semantic analysis, a fundamental bottom-up NLP technique, to the domain of digital games.

Lastly, we avoid a fundamental shortcoming of the work that has been done in this area. None of the previous models can be engaged beyond the publications describing them, which is troublesome given the complexity of machine learning models and the resulting difficulty of adequately describing them. Below, we present not just a model, but a publicly available research tool that itself is a visualization of and an interface to our model. We hope that future research in this area will follow our example of building and releasing tools by which these bottom-up models can be explored.

2.2 Latent Semantic Analysis

Latent semantic analysis (LSA)[1] is a statistical technique by which words are attributed semantic representations according to their contextual distributions across a large collection of text [23]. These computable representations afford direct calculation of how semantically related texts are to one another, which is the fundamental problem in information retrieval, the field in which LSA originated. Though it was specifically developed as a method for automatic indexing and retrieval of documents in large databases [16], LSA became a landmark technique in computational linguistics that has been used in a variety of domains, from literature [34] to science studies [12].

The method is built on the assumption that words with similar meanings will occur in similar contexts and that related texts will be composed of similar words. From a large collection of text, called a *corpus*, a *co-occurrence matrix* of its *terms* (the words and other tokens appearing anywhere in it) and its *documents* (the individual texts it comprises) is constructed. In this matrix, each row represents an individual term and each column an individual document. The cells of the matrix are populated with *frequency counts*, such that each cell will have a count of the number of times the term of the corresponding row occurred in the document of the corresponding column. Since this matrix representation only takes into account term-document co-occurrence, word order in the documents is ignored—*i.e.*, each document is represented as a *bag of words*. Rather than work with the raw term frequencies, however, the cell counts in the term-document matrix are typically transformed. The weighting scheme conventionally used for this purpose is *term frequency−inverse document frequency* (tf−idf) [48], which penalizes terms for appearing in many documents and rewards them for appearing in few.

The matrix at this point can be thought of as specifying a tf−idf *vector space* [49], in which each document is represented as a tf−idf vector (its column in the matrix). Each document's vector will be composed of tf−idf values for each term that occurs in that document and zeros for each term that does not. In a matrix of n terms by m documents, document vectors will thus be n-dimensional. Given the number of terms appearing in a typical corpus, these are likely to be *very* high-dimensional vectors, comprising tens of thousands of entries. The hallmark of LSA is that it reduces the dimensionality of these vectors by a variation of factor analysis called *singular-value decomposition* (SVD) [18]. SVD is invoked with a hyperparameter k, which specifies the desired number of dimensions. It is crucial—and often difficult, as we discuss in the next section—to specify an appropriate number of dimensions for SVD [8]; typically, around 300 are chosen. Once the n x m matrix is submitted to SVD, the k dimensions with the largest singular values—*i.e.*, the dimensions that capture the greatest variance in the original matrix—are retained, with the remainder being set to 0. Put more simply, SVD reduces the number of rows in the matrix while trying to maintain statistical relationships present among the columns in the full matrix.

LSA's use of SVD causes the n-dimensional document vectors to become k-dimensional vectors in the space derived by the SVD—this makes computation more efficient, but more importantly, it allows the model to infer semantic associations that are not encoded in the full tf−idf matrix. This is by virtue of the reduction in the number of rows, which does not cause some terms to be altogether ignored, but rather causes a sort of fusing together of groups of terms that have similar statistical associations with documents in the corpus. The result is that LSA may be able to infer that two terms that do not appear together in any document—perhaps dialectal variants that denote the same thing, like 'gas' and 'petrol'—are in fact highly semantically related [23]. By the same token, it may infer the semantic relatedness of two documents that have no terms in common. This ability to learn global associations from local co-occurrences is the achievement of LSA and the reason that it has found such widespread use. (For argument that LSA instantiates a cognitive theory of human learning, see [23].)

Semantic relatedness between documents is typically calculated by taking the cosine between the documents' k-dimensional LSA vectors. If this is not intuitive, try conceiving of an LSA model as a k-dimensional space in which each document is placed at its k-dimensional coordinates. In this space, the semantic relatedness of two documents is reified as the distance between the documents' positions in the space—this distance is what the cosine represents. In

[1] LSA is sometimes also called latent semantic indexing.

corpora in which each document pertains to a specific individual concept, such as a corpus comprising encyclopedia entries, these relatedness scores can reasonably be utilized as a measure of the relatedness of the concepts themselves. As we explain below, this is how our tool, GameNet, reasons about game relatedness.

Here, it is useful to briefly explain the difference between *semantic similarity* and *semantic relatedness*, which are distinct, though often confused, concepts in computational linguistics [10]. Semantic similarity is a special case of the more general notion of semantic relatedness, which is to say that all concepts that are semantically similar are also semantically related, but not vice versa. As an illustrative example, *mouse* and *rat* are semantically similar (and thus also semantically related), whereas *mouse* and *cheese* are (only) semantically related. LSA homes in on associations between semantically related concepts—which thus subsumes, but is not restricted to, associations among semantically similar concepts—and so in this paper we refer to relatedness and not similarity. Later, we return to this notion to discuss the difference between game relatedness and game similarity.

3. METHODS

GameNet is underpinned by an LSA model trained on a corpus comprising Wikipedia articles for nearly 12,000 digital games. This model represents the first application of latent semantic analysis to this domain. In this section, we describe our corpus and its construction, as well as details surrounding the derivation of the LSA model. Again, we attempt to write for an audience of non-NLP practitioners.

3.1 Corpus Construction

Our corpus is composed of Wikipedia articles for 11,829 digital games. Wikipedia has category pages for each year since the inception of digital games that link to all the Wikipedia articles for games published that year; our corpus was constructed in May 2014 by extracting the text of all the articles linked to from these pages. Initially, close to 17,000 articles were extracted, but we chose to exclude articles that were less than 250 words in length or that were marked as being stubs. Additionally, due to issues during text extraction, a handful of games that do have articles of sufficient length written for them are unfortunately also excluded. Due to our corpus originating in this way, there are many games that are not included in our LSA model and consequently GameNet.

Articles in the corpus, which itself totals nearly 14.5M words, range from 250 to 9858 words in length, with a mean length of 1218 words.[2] We found that a small number of games have articles approaching lengths an order of magnitude above the mean, and that article length generally increases as game year of release becomes more recent. From informal investigation, we observe that Japanese visual novels seem to be especially well-described on Wikipedia, while sports games are generally underspecified. Though it is beyond the scope of this paper, we encourage more rigorous investigation of authoring patterns associated with Wikipedia articles describing individual games.

We *preprocessed* our corpus by removing punctuation and stop words, as well as terms appearing in only a single document, and by lemmatizing all words. Preprocessing is a conventional procedure whereby a corpus is prepared for actual processing by an algorithm like LSA. Punctuation removal is a common step in this procedure because punctuation symbols do not have semantic content in a bag-of-words representation. Similarly, *stop words*, which are extremely common and often grammatically functional words—*e.g.*, ‘the’ or ‘it’—are a classic source of noise for tasks such as LSA due to how frequently they occur in the English language. For this reason, they are typically removed during preprocessing. Because LSA is most often used to measure document relatedness, terms that appear in only a single document are conventionally removed due to constituting idiosyncrasies of their documents that do not help to signify semantic overlap with other documents. Finally, *lemmatization* is the conversion of inflected forms of a word to that word’s canonical form, or *lemma*. In English, this means changing plural nouns to their singular forms—*e.g.*, ‘games’ to ‘game’, ‘children’ to ‘child’—inflected verb forms to their base forms—*e.g.*, ‘jumps’ to ‘jump’—and so forth. Lemmatization is done for the same reason that the text of a corpus is converted to lowercase, which is so that all instances of the same term are identical. For this step, we used the WordNet lemmatizer [32] available in the Natural Language Toolkit suite of Python modules [7].

3.2 Model Derivation

Having prepared this corpus of Wikipedia articles, we derived our LSA model by the conventional method outlined in Section 2.2. Using the Python machine-learning toolkit Gensim [45], we constructed a term-document co-occurrence matrix from our corpus, transformed its frequency counts using tf−idf term weighting, and then derived LSA models for every dimensionality k between 2 and 500. At this point, the major task became selecting an optimal dimensionality.

For some time, we puzzled over how to do this in a way that would best serve our bottom-up scholarly approach. In applications where the performance of an LSA model can be directly measured, one may simply select the dimensionality that maximizes performance. Indeed, this is how we selected dimensionality for a related project involving an LSA-fueled game recommender system [47]; that is, we simply chose the dimensionality that maximized system accuracy. In the case of GameNet, however, we were hoping to build a system that would reason about game relatedness independently from any explicit presuppositions about it. While we could have chosen the dimensionality that best agreed with our own notions of game relatedness, this would have undermined a major design goal for the tool, which was to produce a model that would find interesting game relationships that humans—game scholars, even—would not find on their own.

To avoid this pitfall, we had to eschew conventional (and, at times, enticing) notions of model performance (*i.e.*, accuracy). Instead, we tried out some less conventional dimensionality-selection techniques that are used when a model’s performance cannot be directly measured. In the first, a *scree plot* is drawn using the *singular values* generated by LSA’s SVD step. Each of these singular values corresponds to an individual dimension and serves as a measure of how important that dimension is to the LSA space derived by SVD. Put another way, a singular value captures how well the data could be accounted for by that one dimension alone. The scree plot, then, is a bar graph that depicts the singular values for each dimension in decreasing order; if there is a

[2]The longest article is for *Dragon Valor* (1999).

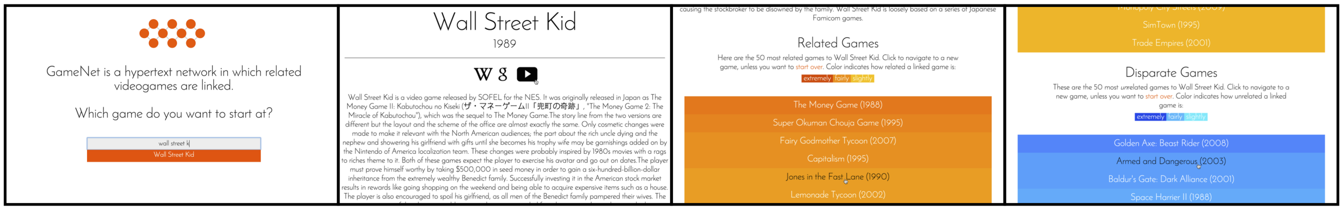

Figure 1: Excerpts from a GameNet exploration of *Wall Street Kid*.

visible elbow in the plot (a sudden drop off between consecutive dimensions), that represents a good dimensionality to select [8]. We carried this out, but found no visible elbow, the existence of which is not guaranteed. However, even if there is no visible elbow in the plot, there may still be one embedded in the data that can be discovered by fairly complex statistical methods (which we will not recount here) [65]; we tried this as well, but again to no avail.

Finally, we settled on another attested, though less objective, method for selecting dimensionality. In this strategy, the optimal dimensionality is the one for which pairs of terms for concepts that have close real-life associations are maximally related [8]. Just like document-document relatedness is calculated by taking the cosine between the two document's LSA vectors, term-term relatedness is measured by taking the cosine between the two term's LSA vectors. The reason this method is less objective than the one described above is that the list of term pairs must be hand-crafted. In applications where the linguistic domain is general English, a typical list would comprise things like synonymous word pairs, *country−capital* pairs, *celebrity−occupation* pairs, etc. For our linguistic domain of digital games, we came up with eleven term pairs for each of the following five pair types: *game−protagonist*, *platform−flagship title*, *game−development studio*, *game−developer*, *sports game−sport represented*. The pair types that we used were conjured rather hastily, we admit, because we had been trying to select a dimensionality for some time (using the unsuccessful approaches we recounted above) and wanted to move ahead with the project. In many machine learning algorithms, there is a model parameter, typically denoted k, that works like model dimensionality does in LSA practice. As we have attested here, picking k is hard! In any event, using this scheme, we selected 207 as our dimensionality, as this was the one for which our 55 term pairs were maximally related.

4. GAMENET

GameNet is a tool for game discovery in the form of a network in which related games are linked. It is intended for use by game scholars (though general game enthusiasts may also find it useful), and is hosted online as a web app—see the link given in Section 7. In this section, we give an overview of the tool and discuss feedback from an expert-evaluation procedure in which six game scholars described their experiences using GameNet.

4.1 Tool Description

GameNet is composed of entries for each of the 11,829 games known to our LSA model. Each game's entry includes links to entries for other games that are related to that game, as well as to gameplay videos and other informative sources found elsewhere on the web. At the GameNet home page (shown in the first panel of Figure 1), the user indicates which game she wishes to start at and is brought to that game's GameNet entry. Here, in a header, the game's title and year of release are prominent, as well as links to the game's Wikipedia article and a YouTube search for *Let's Play* videos of the game (using an autogenerated query).[3] Below this is a summary of the game that was extracted from Wikipedia during the construction of our corpus. The header and summary of the GameNet entry for *Wall Street Kid* [51] are shown in the second panel of Figure 1.

Below these elements is the core of the entry, which is a color-coded listing of the 50 most related games to the game at hand, in terms of their proximity in our LSA space. As alluded to in the previous sections, GameNet judges how related any two games are by calculating the cosine similarity between their documents' LSA vectors. (Because the first dimension in any LSA model is sensitive to document length [21], and because the Wikipedia articles in our corpus are of variable length, we ignore the first entry of all LSA vectors when calculating game relatedness.) On each GameNet entry, related games are listed in decreasing order of relatedness, with background color indicating the degree of relatedness for each related game. To promote exploration, the related games are stylized as hyperlinks to their own GameNet entries. Finally, below the listing of related games is a listing of the most *unrelated* games to the game at hand. These are the games farthest away from it in LSA space and are listed in much the same way, except that the color coding uses cool colors rather than warm colors. This feature is not central to GameNet's intended purpose, but affords teleporting across the LSA space, as it were, where the user may find games, or even genres, that were previously unknown to her. The third and fourth panels of Figure 1 show portions of these segments from the GameNet entry for *Wall Street Kid*.

Here, it is worth returning to the distinction between semantic relatedness and semantic similarity, which we introduced in Section 2.2. (There, we gave the example of *mouse* and *rat* being semantically similar concepts, while *mouse* and *cheese* are semantically related.) From these concepts, in [47] we formally propose *game relatedness*, a more robust notion of game likeness than is represented by conventional genre typologies. Game genres are typically understood as groupings at the level of game mechanics, but two games that are mechanically dissimilar can still be related along several other dimensions. For instance, *Super Mario World* [37] and *Super Mario Kart* [38] belong to distinct genres, but are quite obviously very related games nonetheless. By our notion of game relatedness, all the ways in which two games can be related are all the ways that they could be described similarly. This allows for games to be related according to any notable, shared aspect of their ontologies—

[3]These autogenerated queries use the game's title and platform.

anything that is worth describing about a game may appear in a description of it, and if that same thing appears in another game's description, the two are related. We discuss this here because this is the level at which GameNet reasons about connections between games. While games with similar mechanics are very likely to be connected in GameNet, games may also be connected for having the same designer, for being set in the same fictional universe, or for any other number of characteristics that may be used to describe a game in its Wikipedia article.

4.2 Evaluation

We asked six published game scholars (who had recently conducted studies for which our tool could have conceivably proved helpful) to use GameNet for fifteen minutes and answer a series of questions about the experience.

As a preliminary question, we asked the individuals what scholarly approaches they had employed in their recent projects to research games related to the specific titles or topics they were writing about. Interestingly, though not surprisingly, the scholars listed several methods in total. These included, in no particular order, using Google Scholar and other sources to find related scholarly work; searching Wikipedia for articles describing individual games; playing games using both native hardware and emulation; reading game criticism found online, as well as newspaper articles, magazine reviews, game guides, and game tips that were written at the time of the game's publication (for older games, these included scans and transcriptions and were found across various web sources); watching *Let's Play* videos and other YouTube footage demonstrating speed runs, glitches, walkthroughs, and general gameplay; and, finally, referencing other resources produced by fans, such as walkthroughs and FAQs, as well as a domain-specific informational database (IFDB, the Interactive Fiction Database). We note that the wide variety of approaches these six scholars employed highlights the absence of any single tool for game-studies research that incorporates all the various types of media that they utilized. Interestingly, though, GameNet does include pointers to both Wikipedia articles and *Let's Play* videos, which were each among the enlisted approaches.

Upon answering this initial question, we instructed each of the scholars to start at the GameNet entry for a specific game that was related to his or her recent project. Unfortunately, three of the scholars had hoped to start at games that do not have Wikipedia articles, and which are thus are not included in GameNet (each instead settled on another recent game of study). Our six scholars and the games they started from were as follows: D. Fox Harrell, *Ultima IV: Quest of the Avatar* [42]; Katherine Isbister, *The Sims* [28]; Dylan Lederle-Ensign, *Quake III Arena* [22]; Soraya Murray, *Assassin's Creed III: Liberation* [54]; James Newman, *Super Mario Bros.* [39]; and Aaron A. Reed, *Thomas M. Disch's Amnesia* [14]. (For Harrell and Lederle-Ensign's projects, see [25] and [24]; the rest are currently in submission or still in progress.) Upon reaching the entry for their respective games of interest, the scholars each used the tool for at least fifteen minutes before completing our questionnaire.

We asked whether GameNet would have provided a faster way to locate games related to their recent topics of study, relative to the scholarly approaches they had previously employed. Here, the responses broadly indicated that, as domain experts for their respective topics, they had used the scholarly approaches mentioned above to probe more deeply into specific titles, rather than to seek out additional games related to the topic. Generally, the scholars indicated that, while this would not have helped in their particular recent projects, the tool could prove especially useful as a first method for exploring an area of games that is unfamiliar to the user. "It felt as if it would be more useful to get broad connections in a space I wasn't as familiar in," Reed explained. Isbister, however, appreciated GameNet affirming more tenuous connections between games that she already had in mind. This feeling of being in agreement with the tool on games she already knew led her to be more interested in the games it listed that she did not know about.

When asked whether their fifteen minutes on GameNet led to the discovery of a game that was previously unknown to them or that they had not realized was relevant to their topic, the scholars answered in the affirmative. Lederle-Ensign found multiple titles he had not considered discussing in his study, while Harrell had this to say: "[I came upon] one game I had not thought about much since childhood and seeing it described now made me realize that it had some interesting features relevant to my research." Similarly, Isbister remarked, "I definitely found games that looked promising that I did not know about." Reed, an expert on interactive fiction [44], was surprised to discover an Infocom title he had not known existed. Starting from *Super Mario Bros.*, Newman found three obscure games in Famicom exclusive *Armadillo* [5], Commodore 64 fan sequel *Mario Bros. II* [53], and Wisdom Tree's *Bible Adventures* [55]. "[These] weren't titles I would have got to so quickly, if at all," he remarked. Additionally, Newman was intrigued to find that these games seemed to not be directly related to *Super Mario Bros.*, but more precisely seemed two degrees removed from it by way of Nintendo Game & Watch title *Mario Bros* [36], *Super Mario Bros. 2* [U.S. version] [40], and *Super Mario Bros. 3* [41], respectively. He added, "Getting to games that were similar to ones similar to my original search was quicker with this tool."

As domain experts in the particular areas they explored, Harrell, Isbister, Lederle-Ensign, and Newman all endorsed the connections between games that GameNet listed. Murray and Reed, however, explained that the connections they saw were rather broad relative to their more specific research angles. Interested in finding other titles that took up *Amnesia*'s simulationist approach to interactive fiction—or that, like that title, were authored by a famous fiction writer (in *Amnesia*'s case, this is science fiction writer Thomas M. Disch)—he instead found GameNet's connections to be at the level of genre grouping. That is, the related games he found were merely other examples of text adventures, rather than titles that shared the particular gameplay and production attributes he was interested in. Similarly, Murray was seeking out other games that, like *Assassin's Creed III: Liberation*, have strong female protagonists, but instead found all the other titles from that series (which all have male protagonists) and other games that she felt were related according to broader notions of genre.

Lastly, we requested any additional feedback that the scholars felt like giving. Lederle-Ensign and Reed took this opportunity to praise the interface, and several expressed that GameNet is simply fun to use. "Using it free-associatively (rather than staying based around one core game) is a lot of fun," commented Reed, adding that it is "interesting to see the connection trails it finds." In a similar vein, Newman noted that "there's pleasure in figuring out the connections, particularly as you get further from the original selection." Both, however, wished that GameNet would specifically characterize the nature of the connections it lists, a notion that was central to Murray's feedback as well.

4.3 Discussion and Future Directions

We find several favorable critiques of GameNet from its expert evaluation. Crucially, the connections it makes between games are deemed valid by these scholars in their purview as domain experts. From their feedback, it appears that in its current state GameNet is most useful for cases where the user is exploring a domain of games that she is not expert in. In future work, we plan to evaluate GameNet with regard to this particular use case. In the case that the user is navigating an area in which she is already expert, the evaluation indicates that our tool will at the very least provide additional games that she may not have been aware of or may not have initially considered as being related to her research topic. This highlights the power of bottom-up scholarly methods to yield unexpected findings that are not contingent on any preconceived notions a scholar may be starting from. Furthermore, the fact that each of the scholars did find a new game related to their topic using GameNet is especially remarkable considering the huge assortment of other sources they had already utilized.

GameNet's immediate practical use aside, issues with the tool were raised in our expert feedback. First, the games included in GameNet are currently limited to those that have Wikipedia articles of sufficient encyclopedic coverage written for them, and this was highlighted in the feedback. While the tool does encompass nearly 12,000 games, there are still many notable titles that are excluded. Some of these exclusions appear to be related to particular genres being underrepresented on Wikipedia, which we pointed out in Section 3.1. At some point, we hope to explore Wikipedia authorship in the domain of digital games more deeply, and as we discuss momentarily, we are already setting out to extract a new corpus from a source other than Wikipedia. Additionally, GameNet currently is not particularly useful for exploration starting from a game that is part of a series, since other games in the series will dominate the listing of related games due to being very similar. We are currently developing a feature by which games in the same series can be filtered out from these listings.

The biggest issue with GameNet in its current state is that is does not reason about connections between games at a level of specificity that a game scholar who is studying a topic for which she is already expert may want to seek out. Multiple experts who used the tool were interested in finding games that shared a specific attribute with the game they were researching, but GameNet returned games that were more broadly related. This problem seems to be a byproduct of our LSA model being trained on Wikipedia text. As Wikipedia articles about games are general ontological descriptions of them, GameNet reasons about games with consideration to potentially all aspects of the games' ontologies. When all of these considerations figure into a single score indicating, for instance, that two games are very related, the resulting GameNet connection may seem loose or even complex. Our immediate next step is train a new LSA model using a corpus of GameFAQs walkthroughs with everything but verbs and common nouns filtered out. While connections according to things like a common designer or fictional universe are interesting, game scholars most often reason about game similarity at the level of mechanics and content (*e.g.*, [4], [26]). Using a domain of text that is descriptive of the gameplay experience rather than the total ontology of a game, we believe the resulting LSA model will reason about games at the level of specificity that game scholars do, but still according to a bottom-up process from which interesting and unexpected connections may emerge.

Finally, multiple evaluators indicated that it would be useful for GameNet to give some indication about the nature of each game connection it lists. This opaqueness of the tool's reasoning is deep-seated and endemic of many machine-learning models more generally. Games get connected in the tool because they have similar values along several of our LSA model's dimensions, but these dimensions themselves are formulae that characterize complex statistical associations and are largely uninterpretable by humans. As a few of the scholars added, there is a certain pleasure in attempting to discover by oneself the nature of GameNet's connections, but that this is sometimes necessary is not conducive to the purpose of the tool. We are currently exploring methods by which we can better help the user to interpret the reasoning behind the connections that GameNet makes. Unfortunately, though, this is no small task and will require design insights that, to our knowledge, have eluded those using similar techniques in other domains.

5. CONCLUSION

The huge accumulation on the web of text about digital games is giving impetus to a bottom-up approach to game-studies research utilizing techniques from natural language processing. In support of this agenda, our contribution here has been threefold: we have presented the first complete review of the growing body of work through which this approach has been innovated; we have developed a latent semantic analysis model that represents the first application of that technique to the domain of digital games; and finally, unlike earlier projects that have only written about their models, we have built and evaluated a tool that serves as an interactive visualization of ours. Moreover, we have demonstrated that, beyond being an interface to our model, GameNet may be used more generally as a research tool for game scholars. From an expert evaluation, we find that it is especially useful for the scholar who wishes to explore a relatively unfamiliar area of games, but that it may also be used to discover unforeseen cases related to topics that have already been thoroughly researched. Above all, we hope that this paper will spur future work adopting this emerging methodology.

6. ACKNOWLEDGMENTS

We kindly thank D. Fox Harrell, Katherine Isbister, Dylan Lederle-Ensign, Soraya Murray, James Newman, and Aaron A. Reed for generously providing their thoughtful feedback. This project was made possible in part by Institute of Museum and Library Services grant LG-06-13-0205-13.

7. LINKS

GameNet is hosted online as a web app. Try it out at `http://gamecip-projects.soe.ucsc.edu/gamenet`.

8. REFERENCES

[1] Gamespot. `http://www.gamespot.com`.

[2] Gamestop.com. `http://www.gamestop.com`.

[3] Ign. `http://www.ign.com`.

[4] E. Aarseth, S. M. Smedstad, and L. Sunnanå. A multi-dimensional typology of games. In *Proc. DiGRA*, 2003.

[5] AIM. *Armadillo*. IGS, 1991.

[6] J. S. Armstrong. Unintelligible management research and academic prestige. *Interfaces*, 10(2), 1980.

[7] S. Bird. Nltk: the natural language toolkit. In *Proc. COLING/ACL*, 2006.

[8] R. B. Bradford. An empirical study of required dimensionality for large-scale latent semantic indexing applications. In *Proc. Information and Knowledge Management*, 2008.

[9] J. Brooke and M. Hurst. Patterns in the stream: Exploring the interaction of polarity, topic, and discourse in a large opinion corpus. In *Proc. Topic-Sentiment Analysis for Mass Opinion*, 2009.

[10] A. Budanitsky and G. Hirst. Evaluating wordnet-based measures of lexical semantic relatedness. *Computational Linguistics*, 32(1), 2006.

[11] L. Catalá, V. Julián, and J.-A. Gil-Gómez. A cbr-based game recommender for rehabilitation videogames in social networks. In *Proc. IDEAL*, 2014.

[12] C. Chen and L. Carr. Trailblazing the literature of hypertext: Author co-citation analysis (1989–1998). In *Proc. Hypertext and Hypermedia*, 1999.

[13] C. Chiu, R. Sung, Y. Chen, and C. Hsiao. App review analytics of free games listed on google play. 2013.

[14] Cognetics Corporation. *Thomas M. Disch's Amnesia*. Electronic Arts, 1986.

[15] M. Coleman and T. Liau. A computer readability formula designed for machine scoring. *Journal of Applied Psychology*, 60(2), 1975.

[16] S. C. Deerwester, S. T. Dumais, T. K. Landauer, G. W. Furnas, and R. A. Harshman. Indexing by latent semantic analysis. *JASIS*, 41(6), 1990.

[17] A. Drake, E. Ringger, and D. Ventura. Sentiment regression: Using real-valued scores to summarize overall document sentiment. In *Proc. Semantic Computing*, 2008.

[18] G. H. Golub and C. Reinsch. Singular value decomposition and least squares solutions. *Numerische Mathematik*, 14(5), 1970.

[19] L. D. Grace. A linguistic analysis of mobile games: Verbs and nouns for content estimation. In *Proc. FDG*, 2014.

[20] H. H. Harman. Modern factor analysis. 1960.

[21] X. Hu, Z. Cai, D. Franceschetti, P. Penumatsa, A. Graesser, M. Louwerse, D. McNamara, T. R. Group, et al. Lsa: The first dimension and dimensional weighting. In *Proc. Cognitive Science Society*, 2003.

[22] id Software. *Quake III Arena*. Activision, 1999.

[23] T. K. Landauer and S. T. Dumais. A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104(2), 1997.

[24] D. Lederle-Ensign and N. Wardrip-Fruin. What is strafe jumping? idtech3 and the game engine as software platform. In *Proc. DiGRA*, 2014.

[25] C. Lim and D. F. Harrell. Revealing social identity phenomena in videogames with archetypal analysis. In *Proc. AISB*, 2015.

[26] S. Lundgren and S. Bjork. Game mechanics: Describing computer-augmented games in terms of interaction. In *Proc. TIDSE*, 2003.

[27] J. MacQueen. Some methods for classification and analysis of multivariate observations. 1967.

[28] Maxis. *The Sims*. Electronic Arts, 2000.

[29] R. R. McCrae and P. T. Costa. Validation of the five-factor model of personality across instruments and observers. *Personality and Social Psychology*, 1987.

[30] G. H. McLaughlin. Smog grading: A new readability formula. *Journal of Reading*, 12(8), 1969.

[31] M. Meidl, S. Lytinen, and K. Raison. Using game reviews to recommend games. In *Proc. AIIDE*, 2014.

[32] G. A. Miller. Five papers on wordnet. *Technical Report CLS-Rep-43, Princeton University*, 1993.

[33] G. A. Miller. Wordnet: a lexical database for english. *Communications of the ACM*, 1995.

[34] P. Nakov. Latent semantic analysis for German literature investigation. In *Computational Intelligence: Theory and Applications*. 2001.

[35] D. J. Newman and S. Block. Probabilistic topic decomposition of an eighteenth-century american newspaper. *JASIST*, 57(6), 2006.

[36] Nintendo. *Mario Bros. (Game & Watch)*. Nintendo, 1983.

[37] Nintendo EAD. *Super Mario World*. Nintendo, 1991.

[38] Nintendo EAD. *Super Mario Kart*. Nintendo, 1992.

[39] Nintendo R&D4. *Super Mario Bros.* Nintendo, 1985.

[40] Nintendo R&D4. *Super Mario Bros. 2 [USA Version]*. Nintendo, 1987.

[41] Nintendo R&D4. *Super Mario Bros. 3*. Nintendo, 1990.

[42] Origin Systems. *Ultima IV: Quest of the Avatar*. Origin Systems, 1985.

[43] K. Raison, N. Tomuro, S. Lytinen, and J. P. Zagal. Extraction of user opinions by adjective-context co-clustering for game review texts. In *Proc. Advances in NLP*. 2012.

[44] A. Reed. Creating interactive fiction with Inform 7. 2010.

[45] R. Řehůřek, P. Sojka, et al. Software framework for topic modelling with large corpora. 2010.

[46] P. Resnick and H. R. Varian. Recommender systems. *Communications of the ACM*, 40(3), 1997.

[47] J. O. Ryan, E. Kaltman, T. Hong, M. Mateas, and N. Wardrip-Fruin. People tend to like related games. In *Proc. FDG*, 2015.

[48] G. Salton and M. J. McGill. Introduction to modern information retrieval. 1983.

[49] G. Salton, A. Wong, and C.-S. Yang. A vector space model for automatic indexing. *Communications of the ACM*, 18(11), 1975.

[50] R. Sifa, C. Bauckhage, and A. Drachen. Archetypal game recommender systems. *Proc. KDML-LWA*, 2014.

[51] SOFEL. *Wall Street Kid*. SOFEL, 1990.

[52] C. R. Sugimoto, D. Li, T. G. Russell, S. C. Finlay, and Y. Ding. The shifting sands of disciplinary

development: Analyzing north american library and information science dissertations using latent dirichlet allocation. *JASIST*, 62(1), 2011.

[53] Thundersoft. *Mario Bros. II.* RIFFS, 1987.

[54] Ubisoft Sofia/Milan. *Assassin's Creed III: Liberation.* Ubisoft, 2012.

[55] Wisdom Tree. *Bible Adventures.* Wisdom Tree, 1991.

[56] J. P. Zagal and N. Tomuro. The aesthetics of gameplay: A lexical approach. In *Proc. Academic MindTrek Conference*, 2010.

[57] J. P. Zagal and N. Tomuro. Cultural differences in game appreciation: A study of player game reviews. In *Proc. FDG*, 2013.

[58] J. P. Zagal, N. Tomuro, and A. Shepitsen. Natural language processing in game studies research: An overview. *Simulation & Gaming*, 2011.

[59] M. Zhu and X. Fang. Using lexicons obtained from online reviews to classify computer games. In *Proc. AIS Electronic Library SIGHCI*, 2013.

[60] M. Zhu and X. Fang. Developing playability heuristics for computer games from online reviews. In *Design, User Experience, and Usability: Theories, Methods, and Tools for Designing the User Experience*. 2014.

[61] M. Zhu and X. Fang. Introducing a revised lexical approach to study user experience in game play by analyzing online reviews. In *Proc. Interactive Entertainment*, 2014.

[62] M. Zhu and X. Fang. What nouns and adjectives in online game reviews can tell us about player experience? In *Proc. CHI*, 2014.

[63] M. Zhu and X. Fang. A lexical approach to study computer games and game play experience via online reviews. *International Journal of Human-Computer Interaction*, 2015.

[64] M. Zhu, X. Fang, S. S. Chan, and J. Brzezinski. Building a dictionary of game-descriptive words to study playability. In *Proc. CHI*, 2013.

[65] M. Zhu and A. Ghodsi. Automatic dimensionality selection from the scree plot via the use of profile likelihood. *Computational Statistics & Data Analysis*, 51(2), 2006.