

When the elevator pitch meets the subject heading: how mixtures of other documents can describe what a document is about

Peter Organisciak

Graduate School of Library and Information
Science
University of Illinois at Urbana-Champaign
organis2@illinois.edu

Michael Twidale

Graduate School of Library and Information
Science
University of Illinois at Urbana-Champaign
twidale@illinois.edu

ABSTRACT

We explore the concept of mixture descriptions in the context of film reviews. These descriptions of a film in terms of a combination of two or more other films. This very concrete approach to description can be contrasted with the abstractions typically used in subject headings or the names of genres. By exploring a dataset of film reviews, we uncover some of the features of mixture descriptions as they are used colloquially and investigate when and how they may prove useful. This form of description through combination is not specific to film, and we look at its potential as a bottom-up, ludic form of document description.

Keywords

document description, user-generated content

INTRODUCTIONS

Describing things briefly, clearly and well is hard work. We know that – we study it and try to do it in many parts of Library and Information Science. In this paper, we look at a form of document description that is common in colloquial language but rarely utilized in structured descriptions: description of documents as mixtures of other documents. We study this phenomenon in the context of film, uncovering the patterns of mixture description in film reviews and considering the opportunities and barriers to incorporating it in a formal context. Our aim is to understand how it works – and how it might be used in other settings.

Articulating a movie is challenging – especially if you are trying to do a good job in just a few words. Using a set of nearly 8 million Amazon user reviews of films we find that some people are able to use a very terse and yet surprisingly effective way of describing some aspects of what makes the movie stand out - a qualified mixture of other films. In the

world of mixtures, *Daddy Day Care* becomes “a cross between *Mr. Mom* and *Kindergarten Cop*”, *Looper* is “*12 Monkeys* meets *The Terminator*” and *The Incredibles* plays as “a cross between *Toy Story*, *Superman*, and *Office Space*.” These descriptions seem to get to the heart of the movie, in a way that many people who have seen the movie can agree with. They are clearly inspired by the popular culture view of the movie pitch – where an idea for a movie has to be described to busy executives as clearly and quickly as possible, ideally in an elevator.

We call this *mixture description*. The apparent elegance of this style of communication, when successful, leads us to believe that understanding it may offer ways to describe other types of documents in library collections or archives. We explore this possibility by identifying and quantifying patterns in Amazon user reviews, qualitatively assessing random samples of the data, and comprehensively looking at a case study.

This paper is an initial study of the space of mixture descriptions. This is a potentially large space to study, in areas such as the efficiency of people describing in this way, the satisfaction of receivers in hearing such a description, or the logical challenges of generating mixture descriptions computationally. Here, we provide an overview of how and why mixture descriptions are used, and why the technique is worthy of further focus. While we make this case at times with quantified arguments, this is fundamentally a scoping out of a phenomenon that has implications for describing documents.

We find that those films cited more often in mixture descriptions are chosen for differing qualities, with films within a single description varying along narrative, stylistic, and character themes. The reason that a film is mentioned is not always explicitly stated, yet is clear to a reader. It is this necessary element of interpretation that makes it difficult to quantify and generate automatic mixture descriptions.

Relevance

Mixture descriptions are descriptions of target documents defined by their likeness to mixtures of other documents. The aboutness of a target document is thus explained by selectively adopted properties of the mixture documents.

Though we focus on films, the concept of mixture descriptions is by no means specific to them.

Mixture descriptions are used colloquially in discussing many mediums. We see it in books, artworks, and even cities. For instance, the novel *Twilight* may be viewed as “Jane Austen Meets *Dracula*” (LibraryThing), an artist’s style is “Picasso meets *Yellow Submarine*” (Heller 2008), and Toronto is “New York, run by the Swiss” (Conlin 2005). Such phrases are very lucid for those unfamiliar with the described target, while often offering an ‘a-ha’ element for those that are.

When they work, mixture descriptions are remarkably efficient in their terseness. In Table 1, we took four descriptions from our dataset, and placed them alongside the short film description from the Internet Movie Database. The mixture descriptions are much shorter, but if you are familiar with the films used in the description, they are close to being as informative, if not more. The downside, of course, is the recipient needs to be familiar with the cited films.

Mixture descriptions may be seen as a particularly abstracted entry in the library tradition of subject analysis, which have long focused on thesaurus description such as in subject headings, and more recently has considered facet analysis and social tagging in the face of new information environments (Schwartz 2008). Mixture descriptions are an extension of bottom-up description, a colloquial and uncontrolled lay attempt at aboutness.

While it may not satisfy the desires of library professionals and information scientists for precision and encapsulated

timelessness, there seems to be promise for mixture descriptions in information access. Whereas subject headings improve retrieval, mixture descriptions can communicate the aboutness of a document in an efficient and lucid way.

The greatest flaw of mixture descriptions is that they are context-specific: their efficiency is drastically diminished when the examples in the mixture are not known or paired in a way that does not make the reason for their use apparent. Even here, however, the subjective bias and need for cultural context recalls criticism applied of subject headings, albeit on a finer scale. Critiques argue that subject headings are biased along the lines of gender, religion, ethnicity, and other cultural and personal contexts (Olson 2007, Olson and Schlegl 2001).

Questions

We know how difficult it is to create a good summary of a document or even to say what it is about. Clearly these mixture descriptions are not produced by professional cataloguers. That makes their effectiveness (if indeed they are effective) all the more worthy of note. How do amateur reviewers manage to say something useful about a movie in so few words? In this paper we try to understand the ways that people use different kinds of mixture descriptions.

This involves asking:

- How common are mixture descriptions?
- How do people describe films by the “pitch”: eliciting other films for helping a listener understand a film?
- How *efficient* is it to create these descriptions? What would be lost if we tried to generate them?
- How effective is this form of description for the receiver?
- When do these descriptions fail?

Film	Formal Film Description (via IMDB)	Mixture Description
Daddy Day Care	Two men get laid off and have to become stay-at-home dads when they can't find jobs. This inspires them to open their own day-care center.	“a cross between Mr. Mom and Kindergarten Cop”
Looper	In 2074, when the mob wants to get rid of someone, the target is sent into the past, where a hired gun awaits - someone like Joe - who one day learns the mob wants to 'close the loop' by sending back Joe's future self for assassination.	“12 Monkeys meets The Terminator”
The Incredibles	A family of undercover superheroes, while trying to live the quiet suburban life, are forced into action to save the world.	“a cross between Toy Story, Superman, and Office Space.”
Spacehunter	Three women makes an emergency landing on a planet plagued with a fatal disease, but are captured by dictator Overdog. Adventurer Wolff goes there to rescue them and meets Niki, the only Earthling left from a medical expedition. Combining their talents, they try to rescue the women.	“Mad Max Meets Star Wars”, “Mad Max in space”
King of New York	A former drug lord returns from prison determined to wipe out all his competition and distribute the profits of his operations to New York's poor and lower classes in this stylish and ultra violent modern twist on Robin Hood.	“The Godfather meets Robin Hood”

Table 1: A comparison of select mixture descriptions with their short plot synopses found on the Internet Movie Database (IMDB), demonstrating the terse power of mixture descriptions.

- How does it apply to our ability to communicate information objects in any medium in a clear, understandable way?
- What is it that makes these mixture descriptions effective?
- How might mixture descriptions for films inspire analogous descriptions for other hard to describe kinds of documents?

We offer an initial look to these questions. While we do not claim to have answered them all, we address the initial points on observing mixture descriptions and discuss considerations related to efficiency and effectiveness.

RELATED WORK

This work is a mix of past work on description, combining the more abstract approaches of archetypes, subject classification, facets and folksonomies with the concreteness of item-to-item recommendation.

The study of archetypes naturally lends itself here, as idiomatic themes that serve as a common language of art. Northrop Frye advocated for archetypes of as form of “literary anthropology”, a mechanism by which we can step back from the work and understand the broad strokes guiding its creation (1951). We see the individual parts of mixture metaphors often used in ways akin to archetypes: sometimes a film is more of a symbol than a work. Still, whereas archetypes are a deconstructive activity, mixture descriptions differ in that they are used on a much finer scale with less regard for the inferential consequences of the comparison.

The film pitch approach to describing films has been previously observed by the community at the *TV Tropes* wiki. On the wiki, which is a user-maintained compendium of idioms and tropes in writing, the ‘X Meets Y’ trope collects examples of imagined pitches in the vein of mixture descriptions. Hinting at to the playful nature of mixture descriptions, the page is classified as ‘just for fun’, and the examples that the community creates are not only descriptive but also contains an element of cleverness, as many examples strive for peculiar but surprisingly appropriate comparisons.

A systematized approach akin to mixture descriptions is sometimes done by presenting items within intersections of classification term groupings. The online streaming film service Netflix takes this approach. Netflix offers a browsing model that displays films in unique blends of categories. Similar to mixture descriptions, these categories pair themes across multiple facets, such as style, genre, narrative, and stake-holders. For examples, rather than showing a broad category for dark films or films about show business, Netflix may show a category for “Dark Independent Showbiz Movies.” Research by Madrigal (2014) into these categories suggests the following pattern for mixing themes: “Region + Adjectives + Noun Genre + Based On... + Set In... + From the... + About... + For Age X to Y.”

Netflix’s categories help users understand a grouping of recommended films coherently, but they are also used for

personalization, as part of Netflix’s system strives to recommend the proper mixture of categories for the user (Amatriain and Basilico 2012).

The difference of mixture descriptions from Netflix categories is that the former directly adopts prominent films rather than description terms. That is, rather than saying a film is a “Visually Striking Gritty Film”, one might compare it to “Raging Bull” and leave the rest to the interpretation of the receivers.

A more direct parallel among computational approaches to description is in item-to-item collaborative filtering. Collaborative filtering traditionally performs user-user matches (Resnick et al 1994), so that the habits of a similar User B can inform recommendations for User A. In item-item collaborative filtering, however, recommendations are solely based on items that have been found similar through user activity, such as co-occurring views in a browsing session or products purchased together (Linden et al. 2003, Sarwar et al. 2001). The technique, popularized by the online store Amazon, simply notes when Item A is paired with Item B, serving as a good proxy of similarity.

Mixture descriptions are also comparable to free-text labelling and folksonomies. Folksonomies are inherently colloquial and difficult to control. While numerous different people may think of the same films, the mixture is often a creative act, subject to the context and worldviews of the creator. There does not exist one single ‘correct’ description. Such subjectivity, depending on the setting, can be useful or undesirable. As with folksonomies, it makes it difficult to use the content authoritatively, but it taps into a lay language that matches the needs of many users (Shirky 2005). Weinberger calls this type of loosely-linked classification “third order” information, and argues the variance makes it more informative in the longer term (2007).

Where mixture description deviates from tagging – and many other forms of classification – is in its drastic shift from the abstract. Mixture descriptions are direct and concrete, without sacrificing their interpretive or playful nature.

DATA

Amateur film reviews were our source for colloquial ways of explaining films. We used the dataset previously prepared by McAuley and Leskovec (2013), of 7,911,684 Amazon reviews of films.

The reviews in this dataset spanned 253059 products. Since films can be sold in multiple mediums and editions, the number of films is a smaller subset of the product count.

Focusing on user reviews differs from professional film reviews in two notable ways. First, the reviews are brief: the medium number of words in a review is 101 (ibid). Additionally, the user reviews are contributed in an indefinite timespan. Whereas professional reviews largely are available at the time of a film’s release, amateur reviews can discuss older films. Such time differences are notable,

James Bond	Indiana Jones	Alien	Batman	Rambo
Rocky	Monty Python	Star Wars	Die Hard	Kill Bill
Pulp Fiction	Frankenstein	Predator	Evil Dead	
Dracula	The Matrix	Mad Max	Lord of the Rings	
Matrix	Aliens	Carrie	Braveheart	Bollywood
Beowulf	MTV	Caesar	Wonderland	Hostel
Blair Witch	Blade Runner	Buffy	Blade	Mission
Impossible	Miami Vice	Harry Potter	The Governor	
Godzilla	Scooby Doo	CSI	Blair Witch	Project
Scarface	Pretty Woman	Crouching Tiger	True Lies	

Figure 1: Forty-Five most common terms in “X meets Y” pattern.

because sometimes more recent films are used to describe preceding films as a result.

Our primary preparation of the dataset was to sort and remove duplicate or near-duplicate records. These exist either due to posting errors by the user or quirks of the data. While the sample of reviews is not necessarily complete, removing arbitrary duplicates allowed us to make more reliable comparisons of mixture occurrences relative to the full sample. This cleaning removed 18000 duplicates, less than one percent of the data.

PATTERNS

Mixture Description Structure

In breaking down mixture descriptions, we observe up to three parts: the mixture, a qualification, and a twist.

The mixture cites one or more films that are being comparing the one being discussed. Here we see statements such as, “it’s like *X* and *Y*”, or “a combination of *Y* and *Z*.” Figure 1 notes some common patterns.

The qualification is sometimes paired with a mixture to offer a subjective re-alignment of a listener’s expectations. Qualifiers such as “a better [mixture]” or “[mixture] but without the charm” seems to suggest that mixtures are inexact, and try to correct for when the mixture alone might create a misleading impression.

The twist is another form of modifier that is applied to mixture descriptions, offering a thematic shift from what would be expected by the mixture alone. The twist often modifies style (“a dark...”), themes (“a modern-day...”) or settings (“...in space”).

Not all works are described equally with mixture descriptions. For example, futuristic films and science fiction films seem to be described in this manner disproportionately often, while fewer examples of comedies were seen in our sample.

X meets Y	mashup of X and Y
X / Y / Z	combination of X and Y
mix of X and Y	offspring of X and Y
mixture of X and Y	

Table 2: Common Mixture Patterns

Also, more popular works are not described through mixture comparison as much as more obscure or newer works. This is to be expected, because as works grow more popular they develop their own cultural connotations. We observed this, for example, in comparing the earliest and most recent reviews of the book *The Hunger Games* posted to reading social network LibraryThing. Upon the book’s release, many reviewers noted similarities to other books: it’s “a mix between *The Lottery*, *The Most Dangerous Game*, and Stephen King’s *The Running Man*” (Oct 28, 2008) or “like *Running Man* or *The Lottery*, but updated for our reality show culture” (Dec 4, 2008). In the first 50 reviews, the book is compared to *The Lottery* in three reviews, *The Running Man* three times, *Battle Royale* four times, *The Long Walk* twice, and *The Most Dangerous Game* once. In contrast, the most recent 50 reviews as of April 25, 2014 only have four mentions of other works.

Mixture Types

There are many ways of saying something, a point that is important to remember when dealing with unrestricted description by people (Furnas et al 1987). To get an accurate picture of mixture descriptions in the wild, we need to recognize the most common sentence patterns for comparing a film to other films.

To do so, we developed an initial seed list of possible phrases. Searching through the dataset for these phrases, films that were frequently mentioned were subsequently searched for in order to seek out other ways that mixture comparisons were being made.

Figure 1 notes some of the most common mixture patterns. However, the act of creating them is casual, and the concept of a ‘pattern’ is perhaps misleading. These popular patterns reveal many mixture descriptions, but do not account for all of them.

Films

Table 2 lists the most common films mentioned in “X meets Y” film mixture descriptions.

We see that, at least among the most cited examples, many films function akin to archetypes: a Rosetta Stone for a shared language of film. These films are either representative of a particular genre, or strongly typify a particular visual or narrative style. While occasionally the reason for the similarity is noted, generally the purpose of a citation is unspoken, assumed to be apparent. When mixed, the contexts of what the cited films represent also mixes,

alternating between facets such as genre, themes, and atmosphere.

Qualifications and Twists

Qualifications and twists re-align the expectations of a description if the mixture itself alone would be misleading.

The need for qualifications emphasizes a part of description that is not often conveyed in mixture descriptions: quality. With a few exceptions (e.g. “Plan 9 is the *Citizen Kane* of bad movies”), we generally observed mixtures used to refer to the substance and nuances of a film rather than the quality.

Twists function more like the films in a mixture, representing things like plot devices and style, but usually work on a more general scale. However, they are sometimes interchangeable with more archetypal films. In the *Spacehunter* example presented in Table 1, for example, the film is described alternately as “*Mad Max* in space” and “*Mad Max* meets *Star Wars*.”

Figure 2 shows the terms used often in statements that directly mention a twist. Predominantly, we see occurrences of auteur directors like Quentin Tarantino, Tim Burton, and David Lynch. We see the same pattern when things are described as “X-ian” or “Y-esque” (Figure 3).

Stakeholders such as actors or directors are also sometimes used interchangeably with the works themselves in a mixture description. This is common when they have a distinct modus operandi. For example, we found multiple instances of films described as “Hitchcock meets Tarantino.”

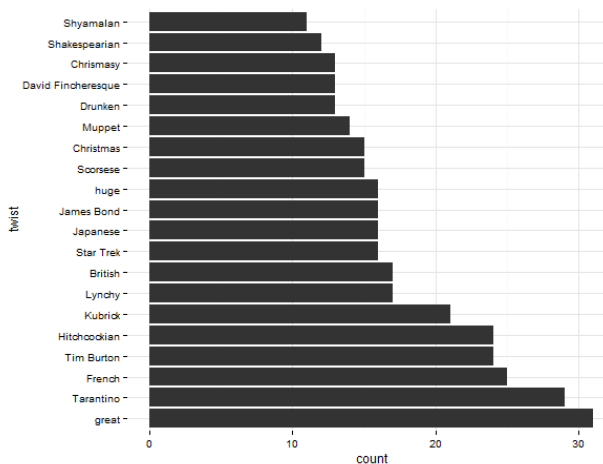


Figure 2: Occurrences of “with a ____ twist” occurrences.

Hitchcockian	Shakespearean	Disneysque
Pythonesque	Orwellian	Burtonesque
Wellesian	Capraesque	Chaplinsque
Felliniesque	Tarantinoesque	Spielbergian

Figure 3: Most common “ian” and “esque” terms.

CASE STUDY

In order to review a population of mixture descriptions, including uses that may have been missed on a broader scale, a close reading was performed for the film *Super 8*.

Super 8 is 2011 Science-Fiction Adventure film, directed by J.J. Abrams. The film is an homage to 70s and 80s Spielberg films in script and in style, causing many reviewers to recall films from that period and genre. Some of the descriptions overlap with the director’s stated influences, others are inferred similarities. The use of mixture descriptions for *Super 8* appears higher than a typical film.

The sample of *Super 8* reviews contained 457 reviews. Within these, a number of films were cited more than once (Figure 4) up to 147 reviews mentioning *E.T.: The Extraterrestrial* (34% of the all reviews).

The films that are frequently mentioned in reviews of *Super 8* stand in for very diverse elements of the film. Films like *The Goonies*, *Stand by Me*, and *The Sandlot* share coming-of-age character themes with *Super 8*. Films such as *Close Encounters of the Third Kind*, *Aliens*, and *E.T.* share narrative themes of hostile or misunderstood extraterrestrials. Meanwhile, films such as *Cloverfield*, *Jaws*, and *Jurassic Park* share stylistic similarities in the directing. Repeatedly, we see the whole spectrum of similarity attributes touched on without specifying the parts of the films that are most comparable to *Super 8*, such as the follow sample review excerpts:

“Wow. ET/Close Encounters of the 3rd Kind/Cloverfield/The Goonies all rolled into one with some Stand by Me thrown in as well.”

“Mixing equal parts of The Goonies, Cloverfield, ET and Red Dawn in the same blender”

“Sort of a mixture of Gremlins, ET, Jurassic Park, the Goonies, Predator, with a little zombie stuff and Dazed and Confused thrown in.”

“E.T. with Jaws”

“Stand By Me/It/Dreamcatcher/ET/Goonies”

“If you liked “The Sandlot”, “Stand by Me” & “Goonies” - this is the movie for you cuz this one rolls all 3 of those movies into one”

A notable portion of the mixture descriptions were also qualified, such as the follow positive and negative qualifications:

“A pretty good mix of “Cloverfield” and “Stand by Me” with a steadier camera and more tense”

“Alien meets Close Encounters, and definitely disappoints”

In Their Own Words

Though it is less common, in some cases reviews would explain their reasons for noting a film. For example, one review notes the character profiles and the narrative themes:

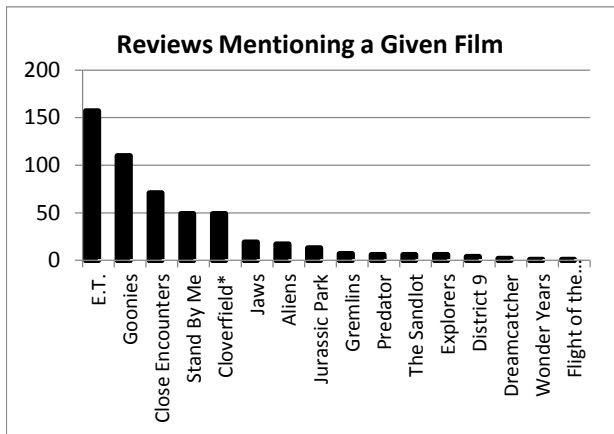


Figure 1: Frequency of films mentioned in reviews of *Super 8*

Coming of age characters reminiscent of *The Sandlot*, *Stand by Me*, and *The Wonder Years* - I grew up with kids just like these.... A story about the military and an alien that is every bit as enjoyable as *Close Encounters* and *ET*.

Another reviewer describes similar films by more specific actions or characteristics:

Take *Goonies* (kids experimenting), *Close Encounters* (the grand evacuation), *ET* (in the end, the alien was a misunderstood cutie), *Transformers* (the self-assembling cubes), *Cloverfield* (the monster is a reduced copy)

A third reviewer notes a more abstract connection related to the quality of execution:

This film also reminds me of *Stand by Me* because it truly captures the mind of a 13 year old perfectly

These explanations are helpful for explicitly explaining the roles of the films being cited, but telling in the fact that they do not provide greatly more information than what is inferred simply by mentioning a film.

DISCUSSION

Indirect Aboutness

Mixture descriptions serve to convey what a document is about, trying to communicate its ‘aboutness’.

Both *Mix* and *And* have been considered as components of aboutness (Bruza et al. 2000). If document *A* is about *X* and document *B* is about *X*, then documents *A* and *B* can be said to be about *X*. Likewise, if document *A* is about *X* and Document *A* is about *Y*, then document *A* can be said to be about *X* and *Y*.

However, aboutness in this view is composed of basic information carriers (IC) – the minimal unit of information. Mixture descriptions use more complex units to describe documents: document *A* may be *like* document *B*, which in

turn may be about ICs *X*, *Y*, and *Z*. This can be seen as indirect aboutness, with a couple of notable consequences.

First, the indirect reference to a property by proxy of another document allows the communication of latent properties. Even if the exact similarity is difficult to formalize in a describable way, the proxy lets one allude to it.

Secondly, while the relation of ‘*A* is like *B* which is about *X*’ suggests transitivity, it is only selectively transitive. That is, some information components of an example are transferred over to a person’s understanding of the document being described, but not all. This is because the relation of document being described to the example documents is one of likeness: a probabilistic rather than objective relationship. While it has been shown that humans hold a transitive reasoning in such probabilistic relationships where true transitivity is not present (von Sydow et al. 2009), it is difficult to anticipate formally.

The transfer of properties seems related to their notability in the context of the document and the context of the pairing. A document can stand in for an information component when it is notably about that component. A cherry tomato can be like a mix of a cherry, by way of its size, and a tomato, by way of its taste. However, even though cherry tomatoes originate in South America and grow on a vine, it would be difficult to comprehend a description of them as a mixture of cocoa beans and watermelons.

It is here where mixture descriptions add an interpretive quality that makes them difficult to formalize logically. Description by proxy and the ability to describe loosely frustrates attempts to systematize mixture descriptions, but we believe that these are precisely the properties that make them appealing to both transmitters and receivers in communication. Indirect aboutness allows a description to recall many different information components in a small amount of space, and the context-dependent nature of which ones are transferred from the examples to also describe the target elicits more imaginative understandings without the need for an extremely eloquent speaker.

Describing Items at Larger Scales

While the terseness and lucidity of mixture descriptions makes them a good candidate for describing items in large information collections, we first need a manner to annotate large numbers of items efficiently

One approach for large-scale annotation is computationally modelling mixture descriptions. As we will argue, however, this is a non-trivial problem, due to the subjectivity and the unspoken subtext of these descriptions, as well as necessary intuition of which films are part of the common language, we believe this to be a non-trivial problem. Instead, because the descriptions are fun and short, crowdsourcing is a more promising method.

One hurdle to building an automated process is that it would need data on similar information objects to the target, but the importance of diversity in mixtures is undermined by many

of the methods for collecting such similarity. Consider, for example, data collection based on item records that are viewed together (as in item-item collaborative filtering) or data collection based on co-occurrence of mentions of the item in written media (e.g. films mentioned together in reviews). These may give us reasonably proxies for similarity, but the similarity does not discriminate in the way we need for mixture descriptions.

Mixtures use items that are similar, but similar in different ways. Put another way, calling *Star Wars* a mixture of its sequels is not as interesting as citing a film about feudal Japan alongside a TV serial about adventuring space explorers.

To further consider this intuition, we applied Latent Dirichlet Allocation (LDA) to train topic models on all the co-occurring film titles in the Amazon user reviews. Subsequently, we built a small testing list of films (*die_hard*, *indiana_jones*, *princess_bride*, *kill_bill*) and inferred the topics that were most likely to have generated those films.

As expected, films that clustered together in topics were not interesting films to group for a mixture. However, pulling prominent example films from orthogonal topics that our target film appears in is more promising. For example, *Indiana Jones* was in two topics that were most strongly represented by *The Mummy* and *Sherlock Holmes*, a curious but somewhat sensible mix. These are connections that are apparent to a human, but such a technique is too noisy to produce mixture descriptions in a clean setting without human intervention. Still, future approaches to automated generation might benefit from similar methods intended to maximize the topical distance between examples that are considered in some way ‘similar’ to a target.

The apparent difficulty of automatically generating mixture descriptions is to be expected: in a way, people use that form of language when they are at a loss for words to describe something. Mixture descriptions often seem to stand in for a *je ne sais quoi* quality, as if to convey understanding while evading explanation.

Thus, while it is not likely that an information system can autonomously start describing its items to users as mixture descriptions, it strikes us that it is an activity where crowdsourcing can lend itself well.

The task seems to be more creative and playful than many existing approaches to crowdsourced description, such as tagging. Like on TVTropes.com, the task of creating mixture descriptions is ‘just for fun.’ In existing systems, we see users tend toward crowdsourcing that is more stimulating than procedural. For example, almost a third of user-contributed content on the social OPAC Bibliocommons is contributions to “Lists” – curated groups of library materials – while tags only account for 1.12% (Spiteri 2011).

How could an information system use crowdsourcing to have their users describe records as mixture descriptions? One observation that we have found in querying students and

colleagues is that mixture descriptions come easily when one is at a loss for words and are quickly understood when seen, but they are very difficult to generate on-demand. If tasking it to an information system’s users, it would likely be more effective to present randomized pairings of items that users can comment on, edit, or rate. It has been observed that in crowdsourcing, users are more likely to respond than to create, compelled to answer when asked about the knowledge or opinions (Organisciak 2010). Even if users have difficulties developing something out of thin air, they are good at responding viscerally and may be inspired by bad mixtures as much as they are delighted by good ones.

This is only one suggested approach, but crowdsourcing of mixture descriptions is a promising direction for an information system provider to explore. More basically, already crowdsourced data can be mined for good mixture descriptions, from places such as Amazon film reviews (McAuley and Leskovec 2013) or LibraryThing book reviews.

Implications

We do not want to simply draw attention to a rather interesting, ingenious (and most likely familiar) practice amongst movie fans. Rather we are interested in mixture descriptions because of what they can tell us about ways to describe other documents that can be complex, nuanced and multifaceted – just as movies are.

As an illustration of the potential of mixture descriptions outside movies, a Google search of the phrase “is like a cross between” yields examples referring to animals, TV shows, clothing, cities, food, vehicles, events, sports, music, musical instruments and people – in just the first 40 results.

Mixture descriptions, when well written, seem to be effective and efficient. They seem to be frequently understandable (but only provided you have seen the other movies being compared), to be able to say something about the movie that the author intends and that many (but not necessarily all) readers would agree with – and all in a remarkably few words.

In LIS we know how hard it is to describe what a document is ‘about’. So it is worth pondering how and why this rather different approach works. It does not have to replace traditional subject descriptors, facets, abstracts, or synopses to be useful.

Typically we describe a document using abstractions – subject headings being a very common resource. However for many more novice users those abstractions can impose an additional learning barrier – they have to learn what the abstractions mean. For an expert in LIS, the technical terms are usefully precise. For others, certain subject descriptions can be rather challenging.

The brevity of the reviews (median 101 words) in the Amazon dataset reminds us that this particular reviewing is done as a leisure activity. It can be viewed as a form of crowdsourced rating and reviewing, as contrasted with that

of professional movie critics. We are also choosing to view these reviews as at least potentially serving as a contribution to crowdsourced abstracting or indexing of the movies. Given the brevity of the reviews we might speculate that the creation of mixture descriptions is actually a way for reviewers to save effort. We still suspect that creating mixture descriptions involves some effort – and indeed considerable creativity. But it might inspire future work to inform the design of crowdsourced alternatives and supplements to cataloguing – an activity widely acknowledged to be very effortful. If the task can be made more playful, it may encourage greater and more sustained participation.

Mixture descriptions seem to be particularly effective when they are ‘unexpected’. That is when they bring together two or more movies that do not seem ‘belong’ together by genre, contentment, or other common measures of similarity. That is, “Star Trek meets Star Wars” seems less effective than “Star Trek meets The Tempest”. The former could be almost any science fiction epic. The latter seems to apply to far fewer possibilities. Carefully picking two very different movies to use in a mixture substantially narrows the number of all possible movies that could reasonably be ‘like’ both of them, thereby increasing the efficiency of the description. Also it can help in getting to the gist of what it is about each comparing movie that is being alluded to in the compared movie.

Mixture descriptions lead us to speculate how and when it might be useful to describe a book in terms of other books – or a dataset in terms of other datasets.

FUTURE DIRECTIONS

As work on mixture descriptions continues, there are numerous fertile areas for explore. These include:

- Satisfaction. Do recipients feel that they understand a film that has been described through a mixture? Do they prefer it over other forms of description? How do they react when a mixture description fails, perhaps through a reference that is too obscure?
- Generation. Can we adopt the findings of this study to describe new films or other information objects?
- Other mediums. Can this approach be useful for things other than movies? Books? Research papers? Datasets?
- Effort and efficiency. How does the effort to create mixture descriptions compare to the effort in understanding them.
- Consistency. How consistent or inconsistent are people in using mixture descriptions?
- Network analysis. By representing works as bound relationships of other works, can we infer second-order relationships and similarities?

CONCLUSIONS

Mixture descriptions are readily recognizable. They are commonly used as informal ways to describe an unfamiliar item, such as a movie that a receiver has not seen; in terms of other movies that the communicator hopes she has seen. In just a few words they can convey a lot of information – provided that the recipient does indeed know the comparing examples. They have a playful aspect that may encourage people to go to the effort of creating them. As such, we believe that a richer understanding of them can provide a friendlier way to describe resources. They seem to be particularly effective at describing unusual movies - those that are difficult describe because they do not fit neatly into a particular genre. Although this initial study has looked at examples in movie reviews, we have many examples of their use to describe other resource. This encourages us to pursue this work to see how they might be deployed in other settings where there are challenges in describing what a resource is about.

REFERENCES

- Amatriain, X., & Basilico, J. (2012, April 6). The Netflix Tech Blog: Netflix Recommendations: Beyond the 5 stars (Part 1). *The Netflix Tech Blog*. Retrieved from <http://techblog.netflix.com/2012/04/netflix-recommendations-beyond-5-stars.html>
- Bruza, P. D., D. W. Song, and K. F. Wong. 2000. “Aboutness from a Commonsense Perspective.” *Journal of the American Society for Information Science* 51 (12): 1090–1105.
- Conlin, J. (2005, September 11). Toronto. *The New York Times*.
- Frye, N. (1951). The Archetypes of Literature. *The Kenyon Review*, 13(1), 92–110.
- Furnas, G. W., Landauer, T. K., Gomez, L. M., & Dumais, S. T. (1987). The vocabulary problem in human-system communication. *Communications of the ACM*, 30(11), 964–971. doi:10.1145/32206.32212
- Heller, R. B. S. (2008, July 27). The Artist at War. *The New York Times*.
- JustForFun/X Meets Y. (n.d.). *Television Tropes & Idioms*. Retrieved April 29, 2014, from <http://tvtropes.org/pmwiki/pmwiki.php/JustForFun/XMeetsY?from=Main.XMeetsY>
- Linden, G., Smith, B., & York, J. (2003). Amazon.com recommendations: item-to-item collaborative filtering. *IEEE Internet Computing*, 7(1), 76–80. doi:10.1109/MIC.2003.1167344
- Madrigal, A. C. (2014, January 2). How Netflix Reverse Engineered Hollywood. *The Atlantic*. Retrieved April 28, 2014, from <http://www.theatlantic.com/technology/archive/2014/01/how-netflix-reverse-engineered-hollywood/282679/>
- McAuley, J. J., & Leskovec, J. (2013). From Amateurs to Connoisseurs: Modeling the Evolution of User Expertise

- Through Online Reviews. In *Proceedings of the 22Nd International Conference on World Wide Web* (pp. 897–908). Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee.
- Olson, Hope A. 2007. “How We Construct Subjects: A Feminist Analysis.” *Library Trends* 56 (2): 509–41.
- Olson, Hope A., and Rose Schlegl. 2001. “Standardization, Objectivity, and User Focus: A Meta-Analysis of Subject Access Critiques.” *Cataloging & Classification Quarterly* 32 (2).
- Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J. (1994). GroupLens (pp. 175–186). ACM Press. doi:10.1145/192844.192905
- Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). Item-based Collaborative Filtering Recommendation Algorithms. In *Proceedings of the 10th International Conference on World Wide Web* (pp. 285–295). New York, NY, USA: ACM. doi:10.1145/371920.372071
- Shirky, C. (2005). Ontology is Overrated: Categories, Links, and Tags. *Clay Shirky's Writings About the Internet*. Retrieved from http://www.shirky.com/writings/ontology_ouerrated.html
- Spiteri, L. F. (2011). Social discovery tools: Cataloguing meets user convenience. In *Proceedings from North American Symposium on Knowledge Organization* (Vol. 3).
- Schwartz, Candy. 2008. “Thesauri and Facets and Tags, Oh My! A Look at Three Decades in Subject Analysis.” *Library Trends* 56 (4): 830–42. doi:10.1353/lib.0.0014.
- The Hunger Games: Reviews. (2008). *LibraryThing*. Retrieved April 25, 2014, from <http://www.librarything.com/work/4979986/reviews>
- Von Sydow, Momme, Björn Meder, and York Hagmayer. 2009. “A Transitivity Heuristic of Probabilistic Causal Reasoning.” In *Proceedings of the 31st Annual Conference of the Cognitive Science Society*.
- Weinberger, D. (2008). *Everything Is Miscellaneous: The Power of the New Digital Disorder*. New York: Holt Paperbacks.