

Michael J. Giarlo

LIS 520

May 9th, 2004

The "Indie Index"

Introduction

For my design project, I have decided to design an index of music fitting into the independent rock/pop (more commonly known as "indie") genre. Specifically, my documentary units will be albums produced and released by independent labels in the 1990s. The indie genre was highly influential upon both mainstream popular culture and the musical underground, especially in the early '90s when the so-called "grunge" movement was identified, and commodified, by the media. It would be worthwhile to provide a resource where information about the most representative or prominent albums of this genre, the ones most indicative of the sound and the spirit of the movement, can be disseminated in an organized manner to users, whether they have a specific interest in the '90s indie genre, or are generally curious about what comprised it. That is, the index should be equally useful to those who are seeking a particular bit of information, and those who simply wish to satisfy a curiosity. The Indie Index is currently up and running at <http://www.eden.rutgers.edu/~leftwing/520/>. [Note that I frequently refer to myself in the first person plural in this paper to give the illusion of representing a design and implementation team of indexers for the project.]

Subject Scope

The "Indie Index" is an index of rock and pop albums produced and released by independent labels during the 1990s, music that is often categorized as "indie" rock. Though the criteria for what makes a label "independent" are a bit fuzzy, the following definition provided by WordIQ.com should suffice for the purposes of this paper:

The concept of an independent record label is a record label perceived as operating outside the sphere of the 'major' record labels, which is to say the few record companies which dominate the recorded music industry in the West. The boundaries are often blurred however, not only because some independent record labels - particularly when they are successful - have often been co-opted as subsidiaries of major labels; additionally, successfully functioning independent record labels also rely in part on international licensing deals, and other deals, with major record labels.

Ten high-level categories, with at least five subcategories included, are: Performers, an alphabetical list of all bands and artists whose albums are in the index; Albums, an alphabetical list of all albums in the index; Years, a chronological list from 1990 through 1999; Labels, an alphabetical list of the labels that released the albums; Styles, an alphabetical list of subgenres, e.g. lo-fi, indie pop, post-hardcore, and shoegazer; Label Locations, an alphabetical list of indie record label locations; Band Locations, a list of countries and states from which the bands and artists hail; Media, an alphabetical list of the media on which the albums were released, e.g. compact disc, cassette tape, and vinyl record; Artists, an alphabetical list of both solo artists and members of the bands; and Producers, an alphabetical list of the people responsible for producing these albums, e.g. Phil Ek, Nigel Godrich, and Steve Albini.

User Information Needs

The intended audience of the "Indie Index" is the general public, ranging from users who have a specific interest in the 1990s indie

genre, to those who are generally curious about what it is and how it has evolved. The index should be equally useful to those who are seeking a particular bit of information and those who simply wish to satisfy a curiosity. The goal is to provide general categories which make sense to the majority of the index's users, and to prominently feature explanations for the terms which may be a bit more obscure. I.e. if a user has a general curiosity about what non-popular music was released during the 1990s, (s)he will probably not know what an album described as "lo-fi" means.

Collection Scope

The collection scope of the Indie Index may be broken down as follows.

Creator: Rock and pop bands who released albums through independent labels in the 1990s.

Date: 1991 to 1999.

Place of Publication: Various. We do not impose a limit on the scope of our collection based on place, though we have concentrated on releases from labels in the United States.

Display Medium: The albums are originally in various media, though we are dealing exclusively with Compact Discs. All data is audio, and metadata is displayed via computer monitors.

Form: CD Audio.

Format: CD Audio.

Genres: Music falling into the broad indie rock and indie pop categories.

Periodicity: Monograph is the closest fit.

Rights: Audio data is restricted to those who have purchased it. Metadata is free for display and usage.

Level: No expertise is necessary, though some familiarity with the indie genre and subgenres thereof may be useful to those seeking specific information.

Language: We do not limit our scope based on language, but most of the material is American English.

Bias: Not applicable.

Quality of Writing: We use no qualitative criteria to limit our collection scope, preferring instead to allow users to make their own judgments of quality.

Reliability: Not applicable.

Technical Level: Usage of the Index requires minimal familiarity of technology, namely website navigation and basic search & retrieval modes.

Collection Domain

Since the documentary units of the Indie Index are rather specific, the collection domain is identical to the collection scope. That is, we make three key limitations to our collection of documents, all of which are music albums: **1)** We will only consider albums released on independent labels; **2)** We will only consider albums released between 1991 and 1999, inclusively; and **3)** we will only consider albums falling loosely into the rock and pop genres. Limitation 1 was chosen because of the wealth of information that already exists for albums released on major labels. Limitation 2 was chosen to keep the project manageable, otherwise we would have had to cover nearly forty years of albums and would have also had the rather troublesome task of determining when indie became indie. Limitation 3 was also chosen for manageability,

eliminating the need to cover independent country, jazz, hip-hop/rap, and electronica, which are each extraordinarily broad categories themselves.

Display Medium

The Indie Index is displayed only electronically, via computer monitors on machines connected to the Internet. We believe our decision to design the Index as an electronic, web-based resource has ensured that the widest possible community has access to it. The one limitation to this is that the Index, in its current implementation, is only in the English language. So fluency or familiarity with English is a requirement for making sense of the Index.

Indexable Matter

The indexable matter from each album is limited to metadata, currently. Were we to seek out and obtain rights to encode audio and include it in the Index, we would gladly do so. Unfortunately, most labels are unwilling to give away the products into which they put so much work and resources. Since we are focusing on the music, or information *about* the music to be more precise, we have decided not to include liner notes or cover art as indexable matter, though both would make wonderful corollary indexes. Currently, we are not including a tracklist with the indexable matter, but it would make a very nice fit to do so.

Analysis and Indexing Methods

As there is no text to index, nor any audio data, only metadata is indexed, and it is all done by our Indexer (who is most assuredly human). After metadata has been entered, a computer program adds records to the alphabetical and relational category indexes. Since we could not find a suitably rich metadata standard, and decided to use the minimalist Dublin Core standard with a few alterations, there are relatively few metadata to gather for each record. Thus, we feel that one to three human indexers can handle this load easily, especially considering that our collection scope and domain are finite and rather limited. Metadata for each album are stored in a database from which they may be queried and displayed in an orderly format. We have decided not to use any vocabulary management given the variety and uniqueness of most of the data in our Index, deciding instead to use a small team of consistent Indexers with one Index Editor whose job it is to check for quality. We believe the aforementioned facts and decisions make for an Index which is both easy to use and maintain.

Displayed and Non-displayed Indexes

Users have available to them three options: **1)** a search interface, which queries record surrogates in a Dublin Core database; **2)** a browse interface displaying an alphabetical index; and **3)** a browse interface displaying a relational category index. All indexes are thus displayed to users, and they may run searches against them if they would rather that than browsing through. We feel that providing the users with multiple views of our metadata makes the Indie Index easily usable, as we accommodate the needs of users seeking specific information and those wishing simply to browse along and see where links take them. As there is no full-text index, we did not have to make any decisions

concerning non-displayed indexes.

Exhaustivity of Indexing

It is difficult to determine exhaustivity since there do not exist easy benchmarks for our documentary units. Where one could compare the number of terms chosen for a text document with the number of pages, our documents lack any such comparison. Generally, there will be between twelve and twenty index terms chosen for each album, which we view as rather exhaustive. Hence, we conclude that exhaustivity is high. For an exemplar documentary unit, I have cited Pavement's "Slanted and Enchanted" album, represented here in its internal Dublin Core format:

```
<oai_dc:dc xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
           xmlns:dc="http://purl.org/dc/elements/1.1/">
  <dc:title> Slanted and Enchanted </dc:title>
  <dc:creator> Pavement </dc:creator>
  <dc:description> Lo-Fi; Indie Rock; Noise Pop; Alternative Pop/Rock
    </dc:description>
  <dc:publisher> Matador Records </dc:publisher>
  <dc:date> 1992 </dc:date>
  <dc:source> Compact Disc; CD </dc:source>
  <dc:contributor> Gary Young (Producer) </dc:contributor>
  <dc:contributor> Stephen Malkmus; Scott Kannberg; Gary Young
    (Artist(s)) </dc:contributor>
</oai_dc:dc>
```

For this documentary unit, there are fourteen index terms, though there could easily be ten to fifteen more if we could have located a more

rich and suitable metadata standard for music. Each index term is an access point, accessible through all three interfaces listed above. For instance, the 1992 index term is accessible via "Year -> Early 1990s -> 1992" in the relational category index and via the "#" area within the alphabetical index. Since the exhaustivity is high and almost none - I hesitate to say "none," since the style or subgenre of an album is a subjective determination - of the index terms used are subject to interpretation, we believe that both precision and recall will be high, which enhances usability.

Specificity of Index Terms

The metadata we want to gather for albums is varied and rather granular, so we have decided to enforce a policy of choosing index terms with high thematic specificity. Since we are gathering very specific and predictable types of metadata, and as we have decided to avoid making determinations about subject or aboutness, maintaining an index with our high specificity policy is less demanding than other indexes with a similar policy, where determining subject and aboutness are the main issues, not to mention the highest hurdles. In the example above, there are very few albums which will have Gary Young listed as a producer. Searching for albums with Gary Young as producer will thus pick out around five documents, which is very specific. Also, we are including in our relational category index the names of all the albums, most of which will uniquely pick out the album searched for. As stated in the section on exhaustivity, we believe the Indie Index will have high precision in addition to high recall, which enhances usability.

Classified / Relational Arrangement

One of the displayed indexes is indeed a relational index depicting the major categories of information. The relationships holding between higher-level categories and subcategories are predictable for the most part, all reflecting some sort of hierarchy. For example, the Year category is subdivided into Early 90s, Mid 90s, and Late 90s, each of which is broken down into the corresponding years falling between 1991 and 1999. The Styles category could be the most confusing category in the Indie Index since such categorization is subjective, and the subgenres may not be familiar to those who aren't music fanatics. The Styles category is broken down into broad subcategories such as Indie Rock, Indie Pop, and Experimental, which are further subdivided into more granular subgenres such as Twee Pop and Math Rock. A somewhat different relationship holds between the Label subcategories. We decided that the best middle-level categories for labels would be geographic descriptors to show at a glance where various labels are located. Nevertheless, all category relations are hierarchical in one way or another, which we deemed intuitive enough for a very broad audience. Nearly all of our categories are mutually exclusive. There are two exceptions: **1)** New York, for example, may be listed both under the Band Location and Label Location high-level categories; and **2)** Steve Albini, for example, could be listed under the Performer, Artist, and Producer categories. (The difference between Performer and Artist, is that the latter lists members of bands as individual entities so that their work may be tracked across multiple projects.) Since the Indie Index provides a fielded search, we do not believe the lack of total mutual exclusivity impacts our index in a negative way. What context may be lacking in the browse interfaces may be provided by a properly constructed search query.

Vocabulary Management

Metadata for each album are stored in a database from which they may be queried and displayed in an orderly format. We have decided not to use a controlled vocabulary given the variety and uniqueness of most of the data in our Index, deciding instead to use a small team of consistent Indexers with one Index Editor whose job it is to check for quality. The Index Editor may, at his or her discretion, choose to use very limited vocabulary management to illustrate helpful associative and equivalence relations. The former are perhaps most useful for linking like styles or subgenres, to indicate for example that those interested in Post-Punk music may also be interested in information about the No Wave movement in late 1970s New York. The latter, equivalence relations, are used sparingly to link acronyms to their expanded forms. We have chosen to avoid indexing acronyms as is, preferring instead to spell information out for users. The limited usage of vocabulary management is employed solely as an aid for our users, so again we feel that our index is highly usable.

Syntax and Search Interface

The syntax of both the relational category index and the alphabetical index uses normal, not inverted, word order, especially as it relates to proper names. The exception to this rule is that articles such as "a," "an," and "the" are listed after a comma at the end of an index term. We decided to move articles to the end so that they do not affect the sorting, and so users who are unsure about the article preceding the term(s) in which they are interested needn't fret over whether it exists or not. Additionally high-level categories, subcategories, and index terms are formatted using different size fonts and indentation levels, which makes the relationships between them more distinct. Otherwise, the index syntax is fairly simplistic, as it is

post-coordinated at the time of searching rather than pre-coordinated by the small team of Indexers.

The syntax for the advanced search is equally simple, since we have broken out the search into multiple boxes in which users can search on multiple fields. Users may decide to type Booleans into the search boxes, but we believe most will decide to use the pre-configured Boolean pull-down menus due to their simplicity, since they interactively construct the query without the user needing to know very much at all about Boolean logic. Additionally, we have included support for limited truncation and wildcarding using the "*" and "?" characters, where "*" matches 0 or more characters and "?" matches 0 or 1 characters. A very succinct explanation of how Booleans and truncation work is included right above the advanced search interface. The simple search interface is available via the main page of the Index, and does a simple search over all available metadata.

Alphanumeric Arrangement

Both the alphanumeric and relational category indexes are arranged by symbol, rather than by concept. Terms and categories are sorted automatically by the programming that powers the index using standard computerized word-by-word sort methods, i.e. using the ASCII values of the characters. The exception to this rule is an adjustment to the sort so that capital letters are sorted alongside lower-case letters, or in other words, the sort is case-insensitive (which is a departure from strict ASCII sorting).

Locators to Documents

The type of locators used to connect index terms to documents is the hyperlink, a visible internal locator, since the Indie Index is entirely web-based. Clicking on an hyperlinked index term, such as "1991," brings one to a page listing surrogate records for all the albums released in 1991. From that list, one may click on a surrogate record link to retrieve more information from a full record. Ultimately there are two intermediate steps between index term locators and the documents themselves.

Surrogate Displays / Record Structure

There are two formats in which surrogate records are displayed, both of which are based on the underlying Dublin Core database. The first format is somewhat akin to an APA-style citation, generated from the metadata associated with the records. The second format is a more common listing of metadata in HTML format, rather than the native Dublin Core (DC) XML. Since Dublin Core was not a clean fit, we also do some on-the-fly translation. For instance, we translate the DC "Source" field to "Media," the "Contributor" field into separate but otherwise qualified "Producer" and "Artist" fields using added information in parentheses, and the "Description" field has been hijacked to represent the "Style" field of the Index. Using this method, we can continue using the Dublin Core standard internally without confusing users by displaying the raw DC. Since each surrogate record is rather small, just a few kilobytes, retrieval is about as rapid as one can imagine in the digital world, though this is of course dependent on the bandwidth of the user and assuming a reliable network in-between. There are no other display options, since our goal is to provide as much relevant information as quickly and intuitively as

possible.

Conclusion

Our goal in creating the Indie Index was to provide a wide audience with a comprehensive resource on 1990s indie music, in order to spread knowledge about lesser-known music scenes and ultimately to encourage enthusiasm about it. To achieve this goal, we decided to keep the website simple, well-organized, and as intuitive as we could. To this end, much of our design and organization elements were based on the good examples of other websites and strong principles of information organization. This goal drove all of our decisions, whether it was to offer two kinds of searching - one powerful without being overly complex, the other simple but equally effective - and two kinds of browsing, or to develop our policies for exhaustivity, thematic specificity, and vocabulary management. We believe we have met all of our goals in the design of the Indie Index and are hopeful that our users are pleased with the project.