# Multimedia Content's Brokerage: An Information System based on LeSiM

Ioannis Karydis
*Ionian University, Greece*
*Creative Web Applications P.C., Greece*

Andreas Kanavos
*University of Patras, Greece*

Spyros Sioutas
*University of Patras, Greece*

Markos Avlonitis
*Ionian University, Greece*

Nikos Karacapilidis
*University of Patras, Greece*

## ABSTRACT

*Metadata-based similarity measurement is far from obsolete nowadays, despite research's focus on content and context based information. It allows for aggregating information from textual references, measuring similarity when content is not available, traditional keyword search in search engines, merging results in meta-search engines, etc. Existing similarity measures do not take into consideration neither the unique nature of multimedia's metadata nor the requirements of metadata-based information retrieval of multimedia. Moreover, customised Information Systems are required in order to support expert users' processes in testing the performance of these similarity measures. This work presents a customised for the commonly available author-title multimedia metadata hybrid similarity measure that is experimentally shown to be significantly more effective than baseline measures. In addition, the work presents an architecture and a web-based implementation of an information system for data collection & validation by expert users that allows distributed, binary and scalar results' ground-truth definition for a similarity measurement that can be used in digital content's identification and sales.*

Keywords: Collection & Validation, Expert Input, Information System, Multimedia, Metadata, Similarity Measure, Lexical Similarity, Author-Title.

## INTRODUCTION

Multimedia Information Retrieval (MIR), such as videos, musical content, animation etc, is ubiquitous nowadays. Search engines, be these for information or sales purposes, identify video content pertaining to a query and present results as ready to be consumed in their appropriate mode (Google search engine, 2018; Bing search engine, 2018) while musical content providers mine preferences through social networks and other sources in order to assist implicit musical queries leading to playlists (Last.fm, 2018; Spotify, 2018).

Research related to MIR has long focused on multimedia's content for producing representations on which to perform retrieval related tasks, such as similarity measurement. The advancement and widespread penetration of virtual social networks has provided another source of information that is contextual to the actual content and mostly refers to social networks users' interaction with or related to the multimedia content. Contextual representations have been shown to significantly boost information retrieval related results in an array of scenarios (Melucci, 2008; Karydis, Kermanidis, Sioutas & Iliadis, 2013).

Despite the aforementioned focus on the content and context derived descriptors from multimedia data, metadata also allow for direct interpretation of their respective multimedia content (Hanjalic, Lienhart, Ma & Smith, 2008). Metadata descriptors, for all their shortcomings, when existing and accurate, offer a set of mostly predefined textual descriptors that allow for fast and relatively computational cheap information retrieval. Moreover, existing text information retrieval methods can be used, up to a degree of success with almost no adaptation, alleviating thus the need for customised methods for preliminary results.

Numerous approaches as to the schemata that best describe multimedia content exist (Smith & Schirling, 2006). In almost all approaches, the notion of a very short textual description (title) of the content as well as attribution of the content to its author/performer is common a phenomenon. These two attributes, the title and author, although not of the best discriminative capacity, exhibit adequate representative capability and are assigned to the content, by its author, *ad hoc*. Another approach introduces four general outputs for design science research: constructs, models, methods, and instantiations (March & Smith, 1995). Authors pointed out that design research could contribute to the applicability of Information System (IS) research by facilitating its application to better address the kinds of problems faced by IS practitioners.

Given multimedia's content- and context- based MIR successful research results, the focus to metadata proposed herein may initially sound anachronistic. Nevertheless, this is far from the truth, as metadata-based MIR is still required for a plethora of research and industry related activities, such as: aggregating multimedia information from textual references (e.g. screen-scraping from html pages), measuring similarity when content is not available (e.g. client-side playlist editing without need to stream data or burden the server), traditional keyword search in search engines, merging results in meta-search engines that do not host content due to intellectual property issues, and many more.

Existing methodologies focusing on metadata-based MIR can be broadly separated into two classes based on whether use of supportive to the actual metadata information is used or not. This supportive information is similar to the aforementioned contextual (but not necessarily from social media), requires collection, is usually not objective and although it may enhance the metadata it can also introduce noise (Metzler, Dumais & Meek, 2007). This work focuses on solely the title and author metadata information of the multimedia content (i.e. no use of supportive/contextual information is done) in order to perform similarity measurement.

For the aforementioned similarity/retrieval tasks to take place, an information system for data collection from sources and results' validation by expert users is also a necessity. Experts' distribution in timezones and space as well as the variety of the characteristics retrieval tasks may afford (e.g. binary or scalar type of similarity, numerous competing similarity methods, volume of data collected, collection automation, etc.) make results' management and processing a task not straightforward to be addressed. Designing ensemble artifacts involves dimensions beyond the technological, because they result from the interaction

of design efforts and contextual factors throughout the design process (Gregor & Jones, 2007). Authors state that as the process unfolds, mutated forms of the artifact emerge and can be distinguished at different points in time.
Moreover, this collection and validation system lays the foundations for a broker e-service that allows querying, identification and selection of multimedia content from prominent web sources and accordingly offers support to e-sales of such content.

## Motivation & Contribution

Existing methodologies for similarity measurement using solely the metadata of the multimedia content are generic as to the type of text applied onto. Accordingly, these take into consideration neither the unique nature of multimedia's metadata nor the requirements of metadata-based information retrieval of multimedia. Moreover, given the large databases of multimedia content's providers and the variety of the content's description, potential customers are hindered in performing multi-interface search and comparison of the available content.

Thus, in order to address the aforementioned issues with a similarity measurement for multimedia author-title metadata, this work:

- proposes a hybrid lexical similarity measure for the common author-title metadata of multimedia entities,
- conducts experimentation in order to verify the increased effectiveness of the proposed similarity measure in comparison to existing methods,
- proposes an architecture for an information system for data collection & validation by expert users, as well as comparative examination of methods on said data, and
- presents a web-based implementation of the aforementioned architecture that allows API based data collection & management as well as distributed, binary results' ground-truth definition for a set of competing similarity measures and lays the foundation of a broker e-service to support digital sales of musical content.

# BACKGROUND

One of the key processes in textual similarity measurement is the representation of the text used in order to apply methods and techniques. Three usually assumed categories of representation (Metzler et al., 2007) are the surface, the stemmed and the expanded. The surface refers to the unaltered text itself while the stemmed is a normalisation / generalisation of the text wherein words are reverted to their stems aiming at the removal of the commoner morphological and inflectional endings (Porter, 1980). Finally, the expanded representation utilises external resources in order to enrich the surface/stemmed representation with contextual information. This work focuses on the surface presentation due to the requirements of the problem addressed herein.

## Lexical Similarity

Metzler et al. (Metzler et al., 2007) presented a number of similarity measures for short segments of text. Although their work revolves around overcoming the vocabulary mismatch and contextual information

problem, and thus is outside the scope of our work, they also propose a hybrid similarity measure that utilises the surface representation as well.
Bearing in mind the shortness of the author-title metadata in terms of number of characters, one may assume that short text semantic similarity and sentence similarity methods lend themselves as applicable approaches in order to tackle the problem addressed herein.
Short Text Semantic Similarity (STSM), in contrast to traditional text similarity methods, such as tf-idf cosine-similarity, aims at semantic level matching (Boom, Canneyt, Bohez, Demeester & Dhoedt, 2015). The focus on the semantic level of the short texts' similarity is, apart from a required feature to judge similarity of meaning, also due to the lack of word overlap in the short texts compared since these are, more often than not, free text expressions of humans. In contrast multimedia author-title metadata require exact match for the author part (e.g. "The Doors" group have nothing to do with actual doors and no relation to an imaginary author artist titled "The gates" that feature the same notion) and a degree of flexibility for the title part (e.g. "Episode V: The Empire Strikes Back" should be a correct result in either original and re-mastered versions, usually described in the title in contrast to an episode of the TV-series titled "The Empire Builds Back"). Moreover, the common size of short texts is far lengthier (10-20 words (O'Shea, Bandar, Crockett & McLean, 2008)) than the usual length of approx. 7 words of the concatenated author-title (see Section *Experimental Setup*). Accordingly, STSM methods are not applicable to the problem tackled herein.
In the same manner, sentence similarity methods are also not applicable to the problem tackled herein as their main focus is on sentences' meaning, usually for the purposes of text summarization and machine translation (Li, Hu, Hu, Wang & Zhou, 2009). In contrast, multimedia author-title metadata are clearly not selected/assigned to multimedia content on the grounds of syntactical and/or grammar rules, nor for their meaning conveying capabilities.

## Baseline similarity measures

In order to compare the performance of the proposed similarity measure, this Section formally describes a set of baseline similarity measures. All of these are applied on the surface representation and are invariably lexical, i.e. are matching words / terms between the query *q* and candidate text *c*.
**Exact:** In this case *q* is character per character equal to *c*, i.e. $\sum_{i=1}^{N} diff(q_i, c_i)$ where both *q* and c are of length *N* and the function *diff* is any symmetric distance measurement function for characters. Due to its nature, this similarity measure returns a binary result, indicating whether or not *q* is an exact match of *c*.
**Substring:** This measure relaxes the requirement of holistic exact similarity between *q* and *c*, by allowing a match to be established if *q* is a continuous exact substring of *c*, i.e. *c* is a match for *q* when $\sum_{i=1}^{N} diff(q_i, c_j)$ where *diff* is as previously defined, *q* is of length *N*, *c* is of length *M* with $N \leq M, 1 \leq j \leq M$and $j = i + \alpha$ where $0 \leq \alpha \leq M - 1$. Accordingly, the substring similarity measurement is evidently a generalisation of the exact similarity measurement with $M = N$ and $\alpha = 0$.
It is obvious that the sizes in number of characters of *q* and *c*, |*q*| and |*c*| respectively, greatly affect the possibility of identifying *c* as a match for *q*, especially when |*q*| is very small in relation to |*c*|. Thus, if *q* and *c* only share a single character substring (e.g. |*q*| = *1*), then can labelled a match, a result that intuitively hampers the measurement's performance. Accordingly, this work introduces a secondary condition for *c* to be a match for *q*, and that is the relative length of *q* to *c* that is left as a variable for experimentation purposes.
**Subset:** In this case, *q* and *c* are split into terms and each term of *q* is searched for exact matches with terms of *c*. In detail, *c* is a match for *q* when $q_{terms} \subset c_{terms}$ where $q_{terms}$, $c_{terms}$ is the set of terms for *q* and *c* respectively and the similarity between terms is based on the aforementioned exact similarity measure. Following the same pattern, the subset similarity measure is a generalisation of the substring similarity measure where the requirement for continuity of the substring's terms is alleviated.

As with the case of the substring similarity measure, *q* and *c* may only share a single term to be labelled a match (e.g. $|q_{terms}| = 1$ while $|c_{terms}| \gg 1$), again a case that affects the measurement's performance. Thus, a secondary condition is introduced for *c* to be a match for *q*, and that is the ratio of $|q_{terms}|$ to $|c_{terms}|$ that is left as a variable for experimentation purposes.

## Data Collection, Validation & Brokerage

Over the last years, Information System researchers are rigorously validating their quantitative and positivist instruments. Although novel professional societies have been formed and grown in prominence and new demands have been placed on the field's research and teaching obligations, the issue of rigor in Information System research has persisted throughout all such changes. A survey is proposed in (Boudreau, Gefen & Straub, 2001), where approaches are suggested for reinvigorating the quest for validation in Information System research via content/construct validity, reliability, and manipulation validity.

It is evident that when addressing crowdsourcing-based ground truth data, not only not one correct answer exists, but the correct answers are not a-priori known; thus it is even more difficult to generate golden units or even use distance metrics.

In information retrieval and machine learning, golden standard databases play a crucial role, as these allow comparison of the effectiveness and quality of systems. Depending on the application area, creating large, semantically annotated corpora from scratch is a time and cost consuming activity. Usually experts review the data and perform manual annotations. In this context, authors in (Nowak & Ruger, 2010) explore how much annotations from experts differ from each other, how different sets of annotations influence the ranking of systems and if these annotations can be obtained with a crowdsourcing approach. Along this line of research, authors in (Welinder & Perona, 2010) propose an online algorithm to determine the "ground truth value" of some property in an image from multiple noisy annotations. Their experiments on MTurk show that the quality of annotators varies widely in a continuum from highly skilled to almost random.

Authors in (Lee & Hu, 2012) discuss the viability of crowdsourcing music mood classification judgments using Amazon Mechanical Turk (MTurk) as they compare the mood classification judgments collected for the annual Music Information Retrieval Evaluation eXchange (MIREX) with judgments collected using MTurk.

The development of a ground-truth set for the evaluation of mood-based Music Information Retrieval (MIR) systems is presented in (Hu, Bay & Downie, 2007). Concretely, a dataset derived from Last.fm tags as well as the USPOP audio collection is utilized and in following, K-means clustering method in order to create a simple yet meaningful cluster-based set of high-level mood categories as well as a ground-truth dataset is introduced. Another similar work is discussed in (Typke, Hoed, Nooijer, Wiering & Veltkamp, 2005), where a ground truth is proposed that can be used for evaluating music information retrieval systems. There, a collection for a list of 11 queries was filtered, and authors employed 35 human experts for ranking the remaining incipits by their similarity to the queries in order to establish the ground truth as a result.

Information (or data) brokerage refers to the "bridging process between disconnected others in a network" (Halevy, Halali & Zlatev, 2019). In this context it is addressed as the aggregation / curation process with the end step of dissemination oriented in monetisation. The bird's-eye-view position of the aggregator allows for collection of information from various disconnected sources, while it's processing with advanced capabilities enables the organisation of the collected content in ways the end recipients cannot achieve easily otherwise.

## PROPOSED METHOD

### LeSiM

The proposed methodology is based on the subset similarity measure presented in "Baseline similarity measures" Section but with a number of alterations.

**Symmetric similarity:** Initially, the measure is turned into a symmetric with the calculation of not only *c* being a match for *q* but also *q* being a match for *c*, and then the ratio of matched to total terms $\frac{|q_{terms}^{matched}|}{|q_{terms}^{total}|}$ and $\frac{|c_{terms}^{matched}|}{|c_{terms}^{total}|}$ are combined using an equally weighted average. This process aims at taking into consideration not only the exact matching subset between *q* and *c* but also their relative subset sizes in a manner that is less strict than the ratio of $|q_{terms}|$ to $|c_{terms}|$. Most importantly though, it turns the measure's output from binary to range and can be thus addressed similarly as with the secondary conditions for substring and subset measures.

**Synonymy detection:** In order to address the special characteristics of each multimedia type, a simplistic and targeted notion of synonymy handling is introduced. As the surface representation utilised herein follows the current naming practices of each multimedia domain's, extraneous information is occasionally included in the author-title metadata. A common example is the variable expression for indication of a guest performer in musical content with (a plethora of alternatives of) the "featuring" term or the inclusion of encoding - quality descriptors (e.g. "1080p", "H264", "DD5.1", etc) in videos' titles. Accordingly, for each *q*'s term included in a synonym set, the rest of the synonyms of this set are additionally searched in *c* without affecting the $|q_{terms}^{total}|$. As such synonymy definition is outside the scope of this work, the proposed methodology is ignorant to its origin that could be based on custom pre-definition or even be learned using machine learning techniques. It should be noted that this notion of synonymy is far from the generic semantic similarity described in the Lexical Similarity Section mostly due to its targeted application. In that sense, the proposed synonymy is domain specific and explicitly defined for the purposes of including the domain's notion synonymy and a breadth that avoids noise introduction. In other words, our approach herein focuses on selective term-related semantic similarity in order to address common naming practices of multimedia domains while not affecting the rest of the similarity process.

**Approximate matching:** In contrast to subset similarity measure's function for similarity between terms being based on exact matching, our method uses Levenshtein edit distance that identifies the minimum number of operations required to transform the searched term into the candidate (Levenshtein, 1966). The introduction of approximate matching between terms addresses a very important characteristic of the author-title metadata, that of their accuracy level. Metadata are nowadays mostly assigned to multimedia content by content creators, and thus are as accurate as possible. Nevertheless, existing content's metadata may well originate from the era that such information was mostly user assigned and thus error prone. Moreover, user created/edited content is bound to include user-created metadata, and thus discrepancies, while also, user generated queries are *sui generis* as to their inaccuracy. The degree of Levenshtein distance in order to assume match between terms is herein left as a variable for experimentation purposes.

### Collection & validation by Experts System

Experimentation on data is one of the cornerstones of scientific processes. The availability of data nowadays is abundant but nevertheless, this very abundance, in some cases transforms into disadvantage as it may lead *to the distraction of information consumers in addition to negatively impacting their productivity and decision-making* (Karydi & Karydis, 2014). The proposed system seeks to address this unfortunate situation by supporting some the information curation's key processes, such as the collection,

storage, and customisation of information flows' entities. This is achieved by allowing manual and automated collection of information as well as by assisting the processes of expert users in determining the ground-truth in the task of similarity's retrieval tasks. Accordingly, the proposed system acts as a meta-search evaluation platform for various definitions of similarity measures based on experts' definition of the ground-truth.

What is more, information curation constitutes a process of selecting existing and/or creating, validating, enhancing and distributing information. For a detailed lifecycle model of the curation process, namely the involved key actions, the work in (Higgins, 2008) can be considered. A very important step of the curation procedure in knowledge mining, where new information is produced by the handling of informational entities in such a way that the relation entities that exhibit between them, are extracted.

Furthermore, an essay addressed the issues associated with knowledge contributions in design science research in IT-related fields, including information systems (Gregor & Hevner, 2013). It aims to assist researchers in identifying appropriate ways of consuming and producing knowledge when they are preparing journal or other scholarly works and to assist editors, reviewers, and readers in more easily identifying the degree of contribution made by a research study. The essay also aims at clarifying some perceived confusions in terminology and the types of contributions from Design Science Research (DSR). The focal contribution is the DSR knowledge contribution framework with two dimensions based on the existing state of knowledge in both the problem and solution domains for the research opportunity under study.

The proposed informational system's business requirement is, for a set of queries, the collection of related to the queries resulting information flows from manual or automated sources, the execution of similarity definition processes on the query & data and the evaluation of the quality of the similarity results obtained, based on the experts' ground-truth. The equivalent user requirement is the definition of a (set of) query/ies and the support on the evaluation process of the quality of the results collected in relation to the query.

### *Analysis and Requirements' Definition*

Following the Software Requirements Specification by (Wiegers & Beatty, 2013), the system's specification section is as follows. It should be noted that generic processes (such as user registration, accreditation, role management, etc) are not described herein without any effect to the generalisation of the system, as these have been widely discussed in the literature and, most importantly, outside the focal interest of this work.

## 1. Preprocessing

**Description:** A designated system's user sets (a) one or more queries and (b) one or more sources to address the query/ies for results, as well as (c) one or more similarity measures to test the sources' replies to a query.

**Functional Requirement 1.1 - Setup one or more queries:** The user inserts the query in a specially devised environment. The query is in textual form. The user signifies the completion of the submission process in order to define the boundaries of the query. The system stores the query and returns verification to the user. The process can be repeated as many times as the user wants in order to define more than one queries.

**Functional Requirement 1.2 - Setup one or more sources:** The user inserts the source in a specially devised environment. The information for each source is in textual form and might comprise of a title, a description, an access method and a type of result. The user signifies the completion of the submission process in order to define all information of a source. The system stores the source and returns

verification to the user. The process can be repeated as many times as the user wants in order to define more than one sources.

**Functional Requirement 1.3 - Setup one or more similarity measures:** The user inserts the similarity measure in a specially devised environment. The information for each similarity measure is in textual form and might comprise of a title, a description, input/output types and its source code. The user signifies the completion of the submission process in order to define all information of a similarity measure. The system stores the similarity measure and returns verification to the user. The process can be repeated as many times as the user wants in order to define more than one similarity measures.

## 2. Information Collection

**Description:** A designated system's user initiates the querying process to the sources and for the results collected expert users evaluate their similarity to the query.

**Functional Requirement 2.1 – Sources' querying:** The user initiates the querying process of the sources in a specially devised environment. In this environment the user inputs the combination/assignment of queries to be sent to sources. The user signifies the completion of the submission process in order to define all the combinations of queries. The system stores the queries to be sent and then dispatches the queries to the equivalent sources. Subsequently, the system stores the collected reply results for each query-source combination. The system outputs the completion of the process and might also output aggregated statistics on the collection process and its results.

**Functional Requirement 2.2 – Experts' evaluation:** One or more expert users evaluate the quality of results collected for each query from the sources based on the expertise. For each query and for each query's results from the sources, the users submit their evaluation as to the quality of the result in order to be a match for the query. Users' evaluation can be in quantitative form such as binary (e.g. match or non match), scalar (e.g. any number between 0 and 10), and quantised (e.g. Likert 1 to 5 scale) or even in qualitative form such as a textual description. The system stores the evaluation of the user and outputs an evaluation completion confirmation.

## 3. Similarity Methods' Application & Results

**Description:** A designated system's user initiates the execution of the similarity measures on each query and its results followed by evaluation of the results of the similarity measures process compared to the experts' respective evaluation.

**Functional Requirement 3.1 – Similarity methods' execution:** For each combination of query and result collected for this query from a source, the system evaluates their similarity using all similarity methods submitted. The process uses as input each query, each result and each method. The execution is done in the background based on the specifics of each similarity method, as submitted during the *1.3 Setup one or more similarity measures* functional requirement and the score each combination achieves is stored by the system**.** The output of the process is a completion confirmation and optionally the vector of the results obtained.

**Functional Requirement 3.2 - Similarity methods' evaluation:** Evaluation of the similarity methods' execution results based on the results from the experts leading to query result's classification performance. The process uses as input the *3.1 Similarity methods' execution* functional requirement results and the experts' evaluation of the results obtained for each query of the *2.2 Experts' evaluation* functional requirement. The process can be done with numerous classification accuracy methods (e.g. precision, recall, or combinations thereof such as f1-score, etc) by assuming experts' results as the ground-truth (in case of qualitative experts' evaluations then manual processes or conversions are required). Optionally, normalisation methods could be applied depending on the relative volumes of results collected from the sources for each query, while multi-fold cross validations could also be applied

for similarity measures that include stochasticity. The process' output is the evaluation of the similarity measures based on the aforementioned accuracy methods.

## *Design & Implementation Issues*

In order to present a more clear view of the logical level design of the proposed system, Figure 1 shows the proposed system's context Data Flow Diagram (DFD). The DFD shows the relations between the key parts of the proposed system and their information exchange in a black-box manner, i.e. without description on implementation issues. The key entities shown in the DFD are the circle representing the processes, the rectangle representing the external entities, the directional arrow representing the information flows and finally, the text within parallel lines representing the data stores.
Focusing further on the data stores, Figure 2 shows the Entity Relation Diagram (ERD). ERD's show in a graphical manner the information stored in a data store / database with focus on the entities, their attributes and their relationships. The ERD is explicitly made simple in order to retain the generality of the proposed architecture. It only features the entities query, reply, sources, evaluation and similarity measure with rectangular shapes, their attributes with ellipsoid shapes and relationships with diamond shapes. Connections between entities and relationships can be total (double line) or partial (single line) as far as participation is concerned, while the numbers on top of lines describe the cardinality ratio for a binary relationship, i.e. the maximum number of relationship instances that an entity can participate in.
In order to better present the temporal characteristics of the proposed system, Figure 3 presents the State Transition Diagram (STD) for the key entities of the proposed system: the query, reply and similarity method. In this diagram, states are shown in rectangles, arrows represent transitions and arrows' textual descriptions are shown as a fraction, the numerator of which is the event taking place, while the denominator is the response of the system as a result of the event. The two cycles represent the initial (single line) and final (double line) states.

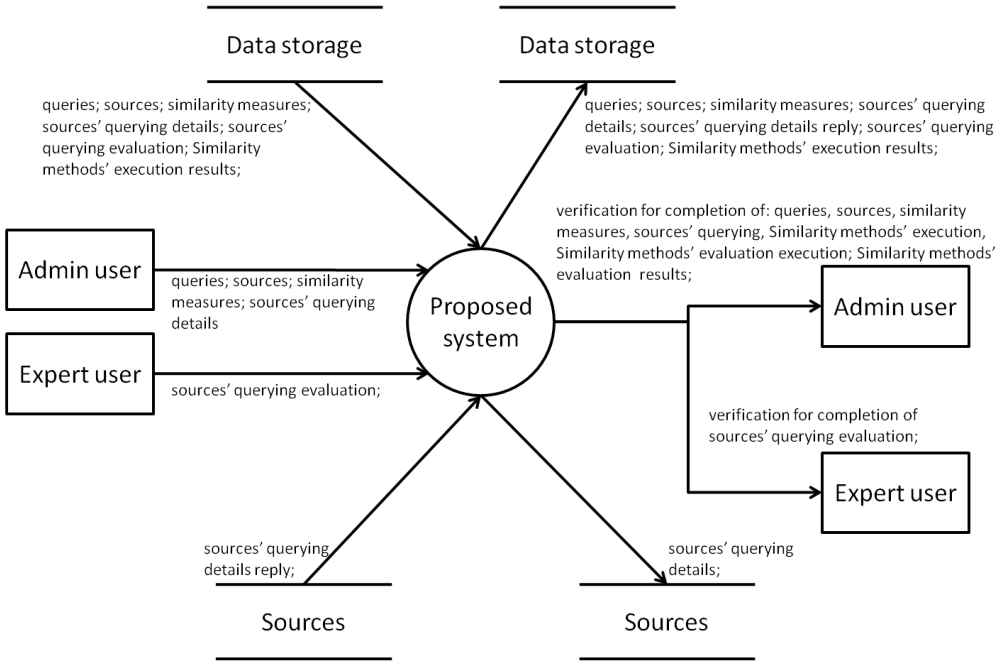


*Figure 1. Context data flow diagram of the proposed Informational System*

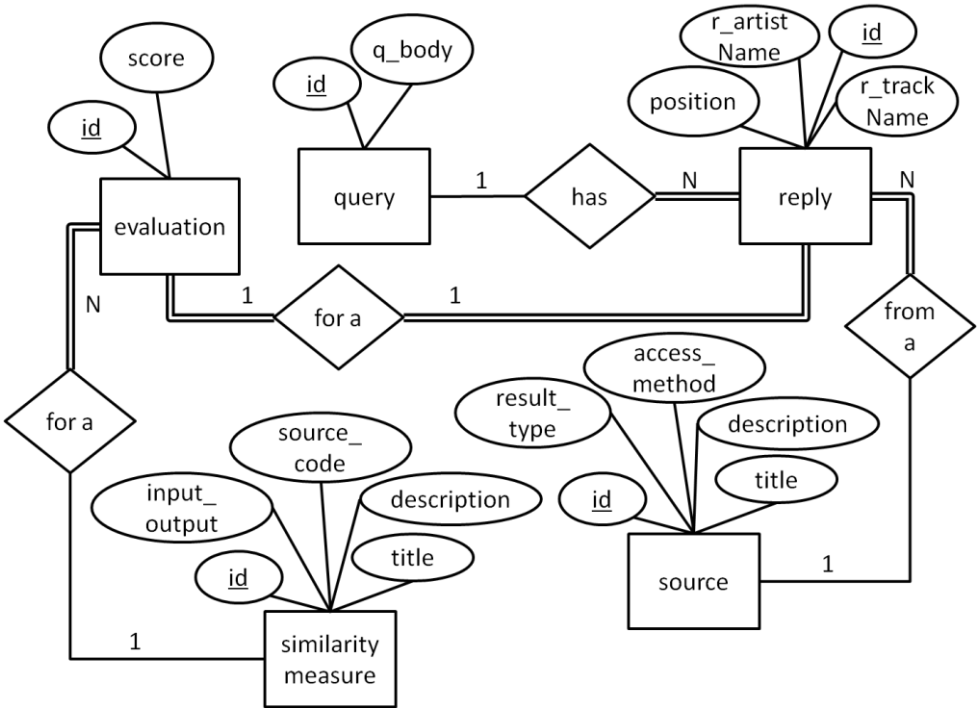


*Figure 2. Entity Relation Diagram of the proposed Informational System*

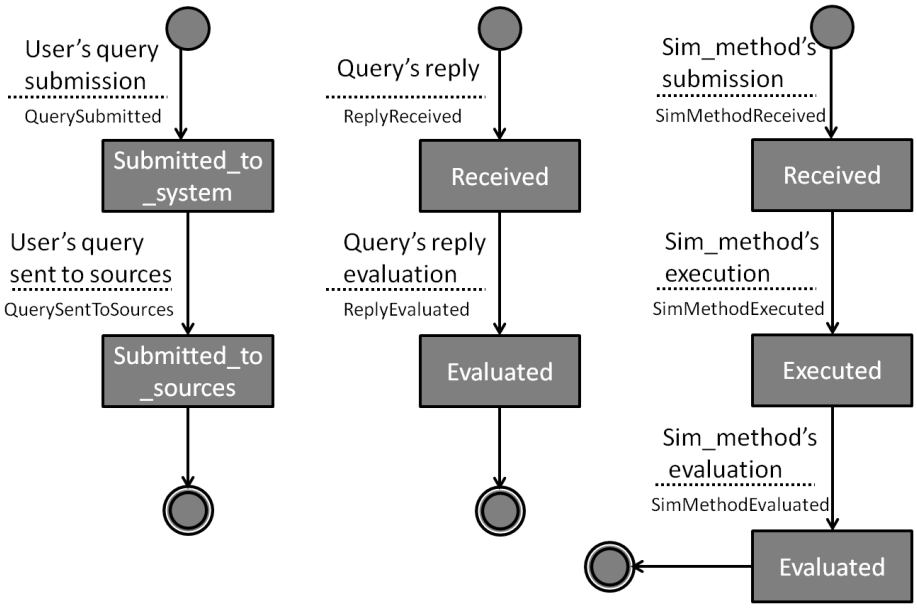


*Figure 3. State Transition Diagram of the proposed Informational System for the entities query, reply and similarity method*

## *The Case of LeSiM*

The implementation of LeSiM is based on the aforementioned generic design with the addition of case specific details. Figure 4 presents the overall architecture of LeSiM.

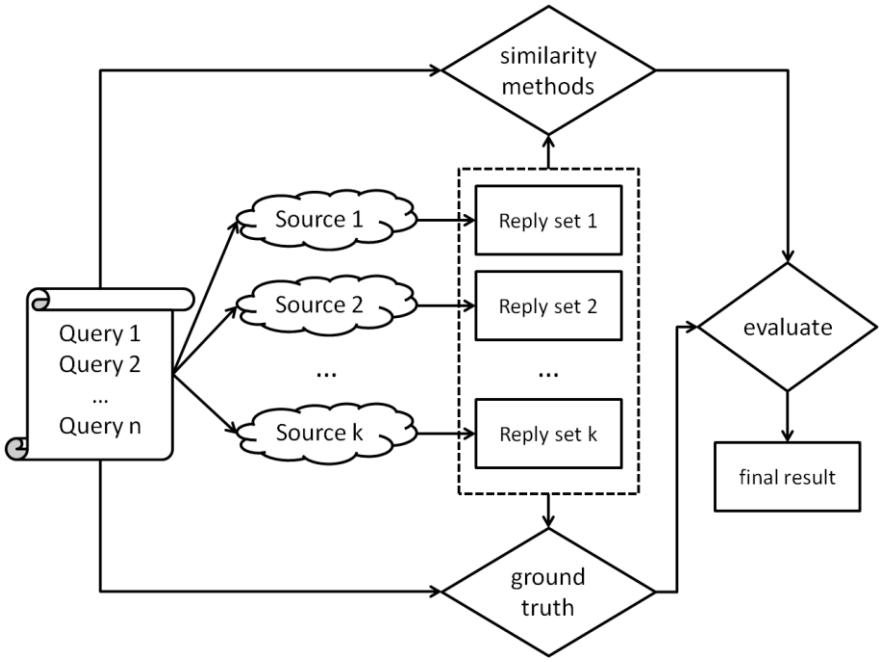


*Figure 4. Architecture of LeSiM*

The most important of the case specific details pertains to its design as a web-service with experts distributed in both temporal and physical collaboration mode. For this to be achieved, the implementation of the database utilised transactional methodologies in order to provide isolation between users accessing the database concurrently. Figure 5 shows the relational schema of the database used LeSiM.

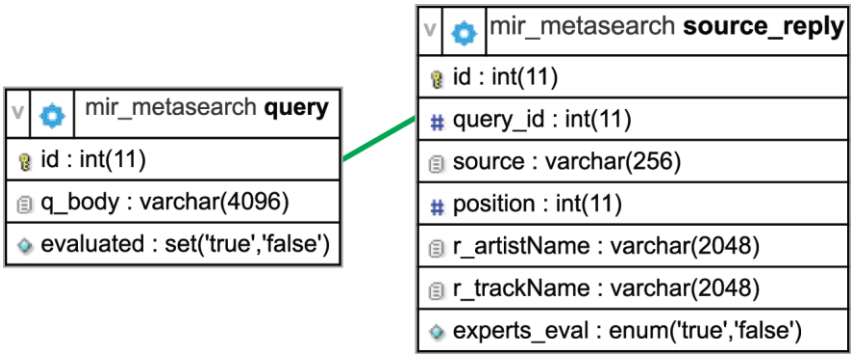


*Figure 5. The relational schema of LeSiM*

LeSiM was implemented in object-oriented PHP scripts, JavaScript for client-side scripting and MySQL for back-end database. Functional requirements 1.1., 1.2, 1.3 and 2.1 were conducted as a pre-processing step, while the functional requirement 2.2 was interactively performed by 3 experts. Finally, functional requirements 3.1 and 3.2 we executed as a post-processing step. Figure 6 presents the interface of LeSiM at the stage *2.2 Experts' evaluation* functional requirement.


*Figure 6. LeSiM's interface for experts' evaluation of replies*

### *Music Brokerage using LeSiM*

The aforementioned implementation of the web-based LeSiM lays the foundation for a musical discovery brokerage service that will: (a) allow users to search more than one providers / e-shops of musical content, (b) provide increased, as shown by the results in the Performance Evaluation Section, discriminative capability in order to furnish increased efficiency search utilities, and (3) is based on an API that is *sui generis* designed to allow accredited access to both functionality and information dissemination for third parties.

The aggregating nature of the web-based LeSiM allows for ease of purchases by clients since the proposed system collects information from various providers, homogenises the presentation of collected data by use of data-warehouse processes and finally presents related results based on the user's query. All in all, such a service can act as an "one-stop-shop" for multimedia content, while also providing alternative channels for content acquisition / purchase.

The effectiveness of the proposed algorithm, addresses the requirement for qualitative results in user information requests. Although the results of LeSiM are indeed a reordered (sub)set of the results of its content providers, LeSiM's reordering is shown to be closer to experts' definition of similarity than the original content providers', allowing thus a better discovery of the available assets.

Finally, as the web-based LeSiM is built on top of an API, it's conversion to functionality and information provider is only natural as APIs allow such activities by machine-to-machine communication. Customised similarity measurement of entities bearing the high-specificity title and author's name can be applied to numerous situations while organised dissemination of the curated metadata information from the aggregated sources to third parties could be invaluable.

## PERFORMANCE EVALUATION

In support of the efficiency of the proposed similarity measure, this Section presents a set of experiments that have been performed. A concise description of the experimentation dataset is also given followed by a performance analysis.

## Experimental Setup

The proposed similarity measure was tested on a dataset of musical content's metadata that despite its focus on one type of multimedia content is generic enough in order to both maintain the generality of the proposed solution as well as function as a proof of concept.
The dataset comprises of 100 queries collected from mining textual information of currently playing tracks from popular Greek radio stations, playing English pop music. Each query was then submitted to 5 musical information providing search engines' API, (iTunes API, 2018; Spotify API, 2018; 7 Digital API, 2018; Last.fm API, 2018; MusicBrainz API, 2018) collecting at maximum 30 results per query per search engine. In total, 9.015 results were collected as a result of the queries, with an average number of 2,1618 terms per author and 5,1138 terms per title. The collected replies' relevance to their originating query was subsequently manually evaluated using a binary classifier.
The experimentation parameters include, for the baseline measures described in the "Baseline similarity measures" Section, the secondary condition ratios for all substring, subset and proposed measures, while solely for the proposed method, the degree of Levenshtein distance. In that sense the secondary condition level acts as the degree of similarity between $q$ and $c$.
The evaluation of the algorithms' results is made by means of precision and recall that was then combined using the F-Measure (Van Rijsbergen, 1979).

## Experimental Results

The first experiment presents the performance of the proposed similarity measure for varying degree of Levenshtein distance and degree of similarity between $q$ and $c$.
Figure 7 shows that the proposed measure is indeed sensitive to the Levenshtein distance, with values of 3 and 4 reaching the best F-Measures, followed by values of 2 & 5 also achieving good performance for the requirement of 100% degree of similarity between $q$ and $c$. The results indicate that the intuition of incorporating approximation for the terms' matching indeed paid-off, evident by the comparison of performance between Levenshtein distance values 0 and 3. Moreover, the results indicate that approximation for the terms' matching should not be thought of as a panacea, since allowing for more approximation, after a certain point, performance degrades, evident by the comparison of performance between Levenshtein distance values 2, 3 and 7. The abrupt change for Levenshtein distance value 1 is attributed to the relatively small size of the dataset under examination.

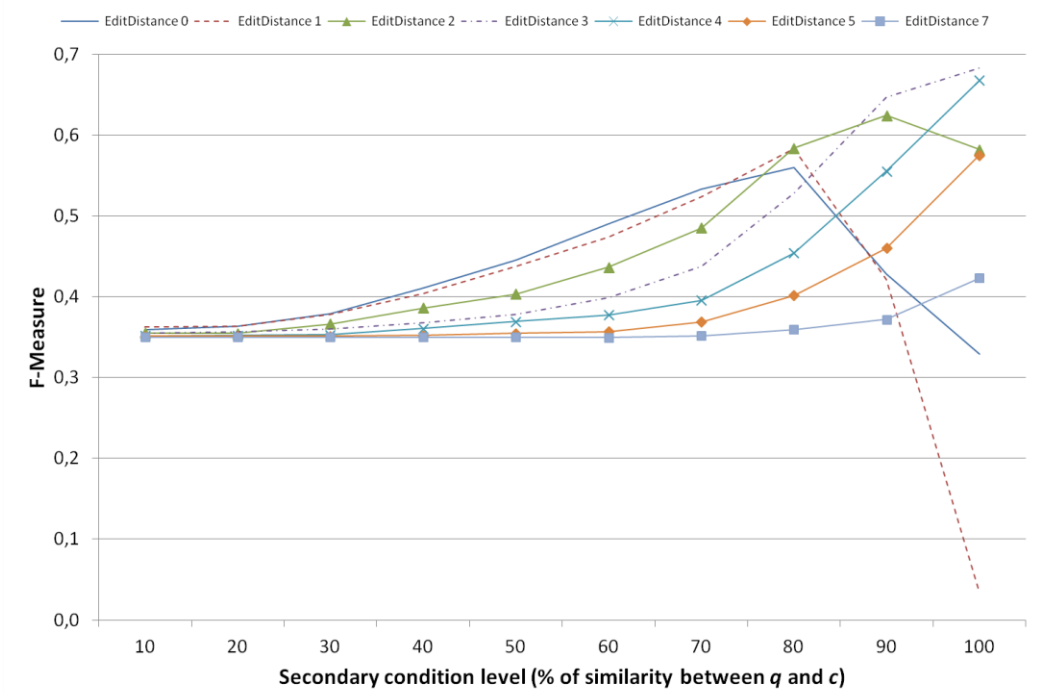


*Figure 7. The performance of the proposed measure for varying degree of Levenshtein distance and degree of similarity between q and c*

The second experiment focuses on the relative performance of the baseline and proposed similarity measures. In this case, the Levenshtein distance for the proposed measure is set to 3 following the best attained result of the previous experiment. Figure 8 shows the attained F-Measure for varying degree of similarity between q and c for all baseline and proposed measures.

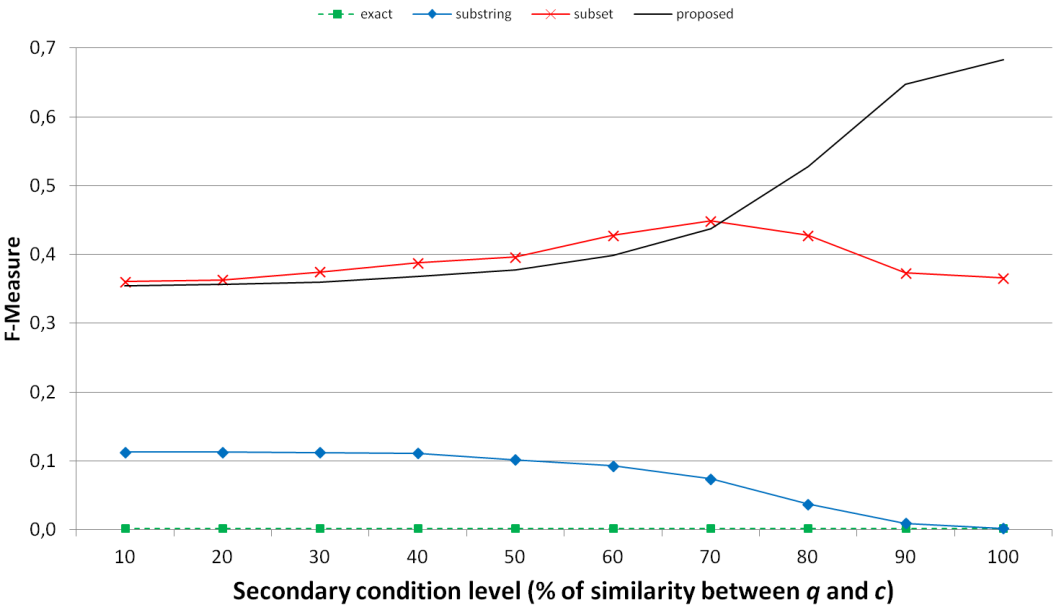


*Figure 8. The performance of all examined similarity measures for varying degree of similarity between q and c*

The results indicate the superior performance of the proposed methodology especially when the requirement of similarity between $q$ and $c$ is in its strictest setting, that is for the substring method is required for the relative length of $q$ to $c$ to be equal to 1, for the subset the ratio of $|q_{terms}|$ to $|c_{terms}|$ to be equal to 1 and for the proposed the weighted average of $\frac{|q_{terms}^{matched}|}{|q_{terms}^{total}|}$ and $\frac{|c_{terms}^{matched}|}{|c_{terms}^{total}|}$ to be equal to 1. It should be noted that the performance of the exact similarity measure is constant as it is not affected by the secondary condition levels.

## CONCLUSIONS

This work examines the use of metadata-based similarity measurement for the purposes of multimedia similarity measurement. Such measurements as usually done on the content or context space with significant accuracy, though a number of tasks are better off measuring multimedia entities' similarity using solely metadata, and especially, the commonly found author-title information. Such tasks include aggregating information from textual references, measuring similarity when content is not available, traditional keyword search in search engines, merging results in meta-search engines and many more research and industry interesting activities. Existing similarity measures do not take into consideration neither the unique nature of multimedia's metadata nor the requirements of metadata-based information retrieval of multimedia.

Accordingly, this work proposes a customised for the commonly available author-title multimedia metadata hybrid similarity measure. The proposed measure draws from the subset similarity measure with significant alterations such as its conversion to symmetric measure, the targeted inclusion of semantic support for specific sets of search terms and the substitution of the exact matching applied during term similarity with approximate matching.

Moreover, the work also presents the architecture for an information system for data collection & validation by expert users, as well as comparative examination of methods on said data. The theoretical architecture is then implemented in a web-based system that allows API based data collection & management as well as distributed, binary results' ground-truth definition for a set of competing similarity measures. The said architecture is also examined as music (case specific) information brokerage system.

Results indicate the superiority of the proposed similarity measure in comparison to baseline approaches, as well as the advantage provided by the proposed customisations/alterations.

Future work includes the expansion of the dataset's size and multimedia content type in order to achieve increased generalisation as well as the re-evaluation of the relevance of results obtained for each query using a likert scale increasing thus the quantisation of the evaluation in order to better map the notion of author-title similarity.


### ACKNOWLEDGMENT

This work is a significantly extended version of the work "Karydis, I., Kanavos, A., Sioutas, S., Avlonitis, M. & Karacapilidis, N. (2018). LeSiM: A Novel Lexical Similarity Measure Technique for Multimedia Information Retrieval. Proceedings Mediterranean Conference on Information Systems".

## REFERENCES


- 7 Digital API. (2018). Retrieved 2018-12-20, from https://developer.7digital.com/
- Bing search engine. (2018). Retrieved 2018-12-20, from https://www.bing.com
- Boom, C. D., Canneyt, S. V., Bohez, S., Demeester, T., & Dhoedt, B. (2015). Learning semantic similarity for very short texts. In 2015 ieee international conference on data mining workshop (icdmw) (p. 1229-1234). Retrieved from http://dx.doi.org/10.1109/ICDMW.2015.86 doi: 10.1109/ICDMW.2015.86
- Boudreau, M.-C., Gefen, D., & Straub, D. W. (2001). Validation in Information Systems Research: A State-of-the-Art Assessment. MIS Quarterly 25(1): 1-16.
- Google search engine. (2018). Retrieved 2018-12-20, from https://www.google.com
- Gregor, S. & Jones, D. (2007). The Anatomy of a Design Theory. Journal of the Association for Information Systems 8(5): Article 19.
- Gregor, S. and Hevner, A. (2013). Positioning and Presenting Design Science Research for Maximum Impact. MIS Quarterly 37(2): 337-355.
- Halevy, N., Halali, E., & Zlatev, J. J. (2019). Brokerage and Brokering: An Integrative Review and Organizing Framework for Third Party Influence. Academy of Management Annals, 13(1), 215-239.
- Hanjalic, A., Lienhart, R., Ma, W.-Y., & Smith, J. R. (2008). The holy grail of multimedia information retrieval: So close or yet so far away? Proceedings of the IEEE, 96(4), 541–547. Retrieved from http://dx.doi.org/10.1109/jproc.2008.916338
- Higgins, S. (2008). The DCC Curation Lifecycle Model. The International Journal of Digital Curation, 1, 3, 134-140.
- Hu, X., Mert, B., & Downie, J. S. (2007). Creating a Simplified Music Mood Classification Ground-Truth Set. Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR), 309-310.
- iTunes API. (2018). Retrieved 2018-12-20, from https://affiliate.itunes.apple.com/resources/documentation/itunes-store-web-service-search-api/
- Karydi, D., & Karydis, I. (2014). Legal issues of aggregating and curating information flows: The case of rss protocol. In *International conference on information law*.
- Karydis, I., Kermanidis, K. L., Sioutas, S., & Iliadis, L. (2013). Comparing content and context based similarity for musical data. Neurocomputing, 107(0), 69 - 76. Retrieved from http://www.sciencedirect.com/science/article/pii/S092523121200759X doi: 10.1016/j.neucom.2012.05.033
- Last.fm. (2018). Retrieved 2018-12-20, from https://www.last.fm
- Last.fm API. (2018). Retrieved 2018-12-20, from https://www.last.fm/api
- Lee, J. H., & Hu, X. (2012). Generating ground truth for music mood classification using mechanical turk. Proceedings of the 12th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL), 129-138.

- Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. In Soviet physics doklady (Vol. 10, pp. 707–710). Retrieved from http://adsabs.harvard.edu/abs/1966SPhD...10..707L
- Li, L., Hu, X., Hu, B. Y., Wang, J., & Zhou, Y. M. (2009). Measuring sentence similarity from different aspects. In 2009 international conference on machine learning and cybernetics (Vol. 4, p. 2244-2249). Retrieved from http://dx.doi.org/10.1109/ICMLC.2009.5212182 doi: 10.1109/ICMLC.2009.5212182
- March, S. & Smith, G. (1995). Design and Natural Science Research on Information Technology. Decision Support Systems 15: 251-266.
- Melucci, M. (2008, June). A basis for information retrieval in context. ACM Transactions on Information Systems, 26(3), 14:1–14:41. Retrieved from http://doi.acm.org/10.1145/1361684.1361687 doi: 10.1145/1361684.1361687
- Metzler, D., Dumais, S., & Meek, C. (2007). Similarity measures for short segments of text. In G. Amati, C. Carpineto, & G. Romano (Eds.), Advances in information retrieval: 29th european conference on ir research, ecir 2007, rome, italy, april 2-5, 2007. proceedings (pp. 16–27). Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved from http://dx.doi.org/10.1007/978-3-540-71496-5_5 doi: 10.1007/978-3-540-71496-5_5
- MusicBrainz API. (2018). Retrieved 2018-12-20, from https://musicbrainz.org/doc/Development/XML Web Service/Version 2/
- Nowak, S., & Ruger, S. M. (2010). How reliable are annotations via crowdsourcing: a study about inter-annotator agreement for multi-label image annotation. Proceedings of the 11th ACM SIGMM International Conference on Multimedia Information Retrieval (MIR), 557-566.
- O'Shea, J., Bandar, Z., Crockett, K., & McLean, D. (2008). A comparative study of two short text semantic similarity measures. In N. T. Nguyen, G. S. Jo, R. J. Howlett, & L. C. Jain (Eds.), Agent and multi-agent systems: Technologies and applications: Second kes international symposium, kes-amsta 2008, incheon, korea, march 26-28, 2008. proceedings (pp. 172–181). Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved from http://dx.doi.org/10.1007/978-3-540-78582-8_18 doi:10.1007/978-3-540-78582-8_18
- Porter, M. (1980). An algorithm for suffix stripping. Program, 14(3), 130-137. Retrieved from http://dx.doi.org/10.1108/eb046814 doi: 10.1108/eb046814
- Smith, J. R., & Schirling, P. (2006). Metadata standards roundup. IEEE MultiMedia, 13(2), 84-88. Retrieved from http://dx.doi.org/10.1109/MMUL.2006.34 doi: 10.1109/MMUL.2006.34
- Spotify. (2018). Retrieved 2018-12-20, from https://www.spotify.com
- Spotify API. (2018). Retrieved 2018-12-20, from https://developer.spotify.com/web-api/
- Typke R., Hoed M., Nooijer J., Wiering F, & Veltkamp R. C. (2005) A Ground Truth For Half A Million Musical Incipits. Journal of Digital Information Management (JDIM), 3(1): 34-38.

- Van Rijsbergen, C. J. (1979). Information retrieval. Butterworth. Retrieved from http://dx.doi.org/10.1002/asi.4630300621
- Welinder, P., & Perona, P. (2010). Online crowdsourcing: Rating annotators and obtaining cost-effective labels. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 25-32.
- Wiegers, K., & Beatty, J. (2013). *Software requirements*. Pearson Education.