**User information needs**

**The Daily Star is a Lebanese newspaper, published in English, which deals with news about a variety of topics covering Lebanon and the Arab world. Plus, it has a small section about international news as well. It is available in print and online. The newspaper covers political, business, social, cultural, sports, entertainment and technology news.** **One fact about the Lebanese education system is that all the materials (math, physics, history, philosophy etc…) are given in English or French and not in Arabic, except for the Arabic language. The type of users targeted is mainly all literate people who want to read the news or a newspaper. So the users are people who know English or have it as a second language & don’t want to read the news in Arabic. Moreover, it targets all the foreign people living and working in Lebanon or in the Arab world who can’t read the news in Arabic. Not to forget, the Lebanese diaspora and expatriates that are living outside Lebanon like for example, in the USA, and are not practicing their native language; therefore, they prefer to read in English. Furthermore, it targets all individuals living outside Lebanon who don’t know Arabic and are interested in Lebanon or the Arab world for instance, many are following the recent development in Lebanon and Syria; an article in the Daily Star will provide a different angle than the one given by foreign newspapers. In the Daily Star users will able to look for recent information as well as historical one, the Daily Star was first established in 1952 but the index goes back only to 1997 we are working on going back all the way to 1952. Indexers are very active to ensure a minimal embargo time. Plus, they will check the index terms twice a year to guarantee efficiency mainly because index terms tends to change, some new emerge and some old fade; this way we can minimize ambiguity and optimize convenience for our users. One other point is that some of the index terms and the way the index is classified can look strange to people that are unfamiliar with Lebanon’s system and culture, for more information please visit the *More info link*.**

**The users will find answers for questions about political events, political leaders and parliament meetings. Plus, users can ask about the stock market and currency or who won in the basketball match and the weather forecast. Not to forget, users can check the movie schedules, art events, plays, and news about the celebrities. Students can use it to search for articles for their political science paper. They can look for articles about everyday life issues for a sociology class. Bank employees can follow the news in the business section. Anybody can use it to search for a specific event. People can read it during their leisure time; they can read their horoscope or play the cross-words. Some look in it for ads related information.**

**Subject scope:**

The subject scope consists of all the facets and aspects of the news that a newspaper can cover. For example, politics and public administration (about Lebanon, Middle East and a small section about international politics), business and economy (in Lebanon, Middle East and internationally), culture and art (festivals, exhibitions, music, obituaries and lifestyle), health (news, reports, diseases, nutrition, medications and tips), environment (in Lebanon and globally), justice and laws (in Lebanon), ads, sports (in Lebanon and globally), media scoops, coverage of events and technology.

**Collection scope:**

**The documentary units consists of articles, reports, editorials, speeches, laws, decree laws, government decisions, weather forecast, ads, obituaries, entertaining games, stock and currency figures and charts, announcements, interviews, photos and videos. The scope of the collection falls under journalism and media anything else is excluded like scholarly works, literary works, children and juvenile material and company reports. Moreover, most of the documentary units should be recent and talks about current topics.**

**The creators of the documentary units are journalists and editorials. The journalists must hold an accredited degree in journalism. They should also have an excellent English language capability, written and spoken. The editorial board consists of professional journalists with a double degree; one in journalism and media and the second in English language. In addition, they must have a good experience in the field of journalism and media. Not to forget, the photographer, the video and audio persons who should at least have a certificate in photography and media production.**

**The newspaper is published daily as a paper copy and online. The documentary units can be viewed online as PDFs as well as HTML. The documentary units of the Daily Star go back to 1952 but the one in the online index go back to 1997. So the documentary units in the index are form 1997 till now. There is a project to go back with the dates in order to cover all the documentary units in the newspaper. It is published in Lebanon and according to the Lebanese copyright law there is no protection or law for the daily news. All daily news will be excluded from the protection in the copyright law (WIPO, 1999). There will be no notes, no bibliography, only the author name will be provided if applicable. All the documentary units are written only in English.**

Usability reflects the users’ needs. Users are able to search and find sources about a variety of news topics. There are some limitations like no documents published before 1997. Plus the time gap for the recent issues to be indexed, however from the Daily Star Home page users can check the today’s issue and go back and check past issues from two weeks. Because they won’t be able to find news from the recent issues using the search option due to the embargo time. Users can search using their vocabulary or they can access the subject list. They can search multiple entities and they can combine many search combinations using our faceted search. The search interface is very easy to use and provides advanced filter options. The intended audience is stated in the [Users Need Section](#User) (clickable hyperlink). The online version is free.

**Collection domain:**

The selectors of the documentary units or the persons who decide which documentary unit to add or not are the editorials. They are professional journalists who hold two accredited degrees, one in journalism and media and the second in English language. Plus, they should have a minimum of ten years’ experience in the field of journalism and media. So the staff must be selected on a high level of professionalism, communication and language capabilities. Our staff members should be able to handle multi-lingual tasks. Additionally, we need professional indexers or taxonomists in order to be able to index the documentary units, so users can easily find what they are searching for, when using our online index. The indexers should hold an accredited degrees and show advanced language skills in English. It is not easy to index newspaper articles especially in a classified index. Every day we have many new articles to be indexed. The indexers will follow an agreed upon rules and guidelines, and teaching new comers is on a whole new level; it is costly and time consuming. The classified index and the subject list will be examined twice a year to make sure that all the terms are still relevant and if a change needs to be done, a collaborative decision shall be taken by our indexers. Not to forget, the archives section where all the documentary units are kept and secured. Large servers are needed; security and a backup must be established. Thus, we need an IT team that is always around. The team will consist of computer scientists and engineers, website architects (to work on our website design and search optimization) and technical staff. As you see the maintainability is very crucial, costly and hard in order to persist and continue.

**Classified/relational arrangement and/or facets**

In the Daily star we always tend to cover and meet all our users’ needs and preferences. In order to do that we will follow a hybrid model using both a classified index and facets with the ability for the users to tag and label the documentary units using their own terms that will better reflect their understanding which will allow them to search using tags. In this way we are providing two options; the top down classification and the bottom up as well. Our users will have the option to choose which model to use. The classified index provides a consistent and meaningful relationship between the headings, subheadings and categories; users will be able to access and browse the classified index from the Home page by clicking on the top level categories in the left side bar. Usually, you cannot be lost, it is a very systematic guide to the information you are looking for. It presents all the concepts in different classes. It is listed hierarchically, and we can go from top to bottom or from broader to narrower. It helps the user to look up the information they want even when they don’t really know what they want. It can be used as a list that can be browsed where users can pick a topic or choose a case or a more specific topic in a bigger field they are studying or looking for. I think that a classified index is way better than an alphabetical arrangement because it gives us more control. However, classified index needs expertise, time and budget. All the headings and subheadings terms are checked on a regular basis, by our taxonomists and indexers, to make sure that they are still in use and still have the same meaning and emphasize the same concept; because language changes, new words emerge, some change and other become obsolete. **Our indexer will update the index while preserving its integrity.**

As for the facets it is newer than the classified index, it is more flexible, topics can be infinite and it can be more expressive. I think it works great with these documentary units because they are all news articles or reports, they are not different. In addition, it is easier to use and it let us combine facets or entities which are mutually exclusive. Users can use index terms that are not pre-coordinated and benefit from post coordination. Users can search using more combinations and options, which can lead to more specific and better search results. The articles can be found in more than one way. Not to forget that, we don’t have to be knowledgeable or professional indexer. Usually, facets works better with users who know what they want.

The facets used are: subject, people (as subject), company (as subject), document type, section, place (as subject), tags and media. In addition to the bibliographic ones (title, author…)

For example, users can build a combination to get business information that is about a politician who is also a business person by using the name of this politician and selecting the business section.

Facets are friendlier and easily used and give the users more freedom and options. Yet, the users should know what they are looking for. Otherwise, I think they should take a look at the classified index and look at the categories, headings and subheadings. The classified index only uses terms that express one concept; there are no pre-coordinated terms. Users can check the terms used and if they want they can combine those using the faceted search. **One other point is that some of the index terms and the way the index is classified can look strange to people that are unfamiliar with Lebanon’s system and culture, please for more information visit the *More info link*.**

**Finally, using tags is the easiest way to search but not the best one or the most effective one. Not to forget, that the tags will be subject to control by our professional index. Users can find a guideline and policy about tagging in the search tips.**

**Vocabulary management, including method for vocabulary control**

The Daily Star index uses a controlled vocabulary list called ***Subject List***. Some of the features offered by the controlled set of terms are:

Scope notes which provide a small definition and history of the term and in what context it is being used. For example, in Lebanon 14 March is a coalition name for multiple political entities that combined on the 14th of March; In Lebanon we refer to this combined group by 14 March; so it is not a date. A small definition for terms like this is crucial and essential to disambiguate and elucidate the information about the subject terms for the users.

Equivalence assistance is provided for example, 14 March or 14th March, harbor or harbour, revolution or revolutions. (Some schools in Lebanon teach the English English and others teach the American English).

Synonymy assistance is provided too for instance, army or military (use military); Lira or the Lebanese Pound (use Lebanese Pound), gas or fuel (use fuel). These examples are visible on the design page 4.

Broader or narrower terms for example, parliamentary elections and elections (election is the broader term), corruption and bribery (corruption is the broader term).

Related terms for example, 8 March see also revolutions; visible on the design page 4.

Acronyms help for example MTC (Mobil Telecommunication Company) or (Mediterranean-Type Climate).

Hints will be provided for users upon the search. The subject list is linked to the search where users will get direct hints to optimize their search quality and lessen their search efforts; not all the users will access the controlled vocabulary list on their own, usually many users go with the rule of the least effort.

Each article is associated with two or more subject headings from the list. Subject headings are arranged in hierarchies so users can choose from broader to narrower and related headings too. Synonyms are included as references so users get to choose the right term. When we click on a subject heading we can see a “scope note” that explain the use of the term, its history and provide reference to related headings so the search will be more clear, less ambiguous and effective. The list is very helpful it can guide the users and put them on the right track when it comes to choose the correct terms for the search. It gives more control for the searcher and renders the search to be more specific and precise. Our team of professional taxonomists and indexers is responsible to maintain and update the subject list. Subject terms and the relationships between the terms are examined periodically and updated to make sure that they are still used correctly, because language changes, new words emerge, some change and other become obsolete.

Moreover, the list will help us to keep track of the used terms along the years in our index. It is a reliable record that we can always go back to. Plus, new indexers will benefit a lot when using it, it is a guide. Language changes over time that is why we need this reliable list. We cannot remember what we were using before to index a certain topic, so we go back to the list and check. To build and maintain a subject list is not easy and very costly. Errors can occur that is why our team works collaboratively and follows a set of agreed upon rule and procedures.

The terms in the list are single and compound; sometimes one word is not enough. The syntax used is the natural language. Names of companies, products and place are used as they are. People’s names are entered as the following family name, first name.

**Analysis and indexing methods**

Human indexers or taxonomists will be employed to do the indexing of the documentary units. These indexers are professional in the field of classification and indexing; in addition, they should demonstrate competence in the fields of journalism, media, news and politics. Similarly, they should have very good language skills in English. The group of indexer will follow an agreed upon list of rules and guidelines in order to maintain consistency. The creators of the documentary units are asked to provide with every document or article an abstract and a set of index terms that best describe it. This process may affect the judgment of the indexer but it will make the work easier and help us gain time in order to minimize the embargo time we need to index the documentary units. All political articles or documents may have similar terms and high occurrence of this terms, this is why we need human indexer in order to understand the concept of the article and differentiate it from the other ones “indexing, as performed by a human being, is a mental process or intellectual analysis, based on understanding the meaning of the text” (Wellisch 1995). The indexers will use the controlled terms when indexing and if there is a need for introducing a new term all the indexers can meet and take a collaborative decision. The list of controlled terms is checked every six month and is subject to update; new words emerge and some old index terms become obsolete. The controlled list was first based on the CSA Worldwide Political Science Abstracts Thesaurus and then edited in order to describe better the Lebanese system and culture. The index terms are chosen according to the literary warrant and the use warrant as well. The index should have a logical, coherent and consistent structure with a high level of standardization, which will allow users to browse and use it easily. “In general, indexes where the semantic relationship between terms is consistent and meaningful are easier for people to use” (Wacholder, 2013). The index follows the principle of collocation, predictability and consistency. It has a meaningful organization; it is classified, hierarchical and relational with cross-reference.

As for the tagging system, users will be able to tag each documentary unit and assign their own tags to it. Tags will be subject to control by the indexers before being posted. We developed a small policy (a form of guidelines) for tagging. Users when tagging will read a brief statement about the synonymy problem, offensive and trashy tags and a link to our subject list (controlled vocabulary).

The cost of producing and maintaining our high level index is really high. No to forget, training new comers and making sure that everyone is abiding by our rules and guidelines. All of this falls under the umbrella of the users’ satisfaction and the usability of the index; our main goal is that our users are able to search, browse and find what they want in an easy way mainly because users always tend to prefer convenience.

**Depth of indexing**

The degree of this index’s specificity and exhaustivity is generally high; the index has high level of exhaustivity and high level of specificity but not too specific “be specific, be specific—don’t be too specific” (Wellisch, 1995). The surrogate records state all the covered themes and the sub-themes as well and cross references if available. Not all the documentary units have the same level of exhaustivity, it depends on each one separately. There are no limitations on the number of the index terms. The index of the Daily Star should be specific and exhaustive because of the subject scope and size of the documentary units, even if a full-text search is available. Many of the documentary units talk about common topics especially in politics yet each may have more than one specific case, so we need a high level of specificity in order to provide the users with a good, precise and effective information retrieval system, while preserving a good recall for retrieving relevant results too. Keeping in mind the scope and size of these documentary units which is really big; the Daily Star is a daily newspaper. This is another reason why the index should have a high degree of specificity and exhaustivity. From the displayed subject terms in the record users will be able to acknowledge the main and sub themes of the articles and understand what it is about. The high level of exhaustivity and the high level of specificity will make a balance between recall and precision.

**Syntax for representation of displayed indexes**

In the displayed index the syntax will be the natural language, which is more common and friendly to the users; probably many of our users are not familiar with the indexing process. No double entries will be entered; instead cross references will be used. The headings are mainly single words or compound words. Some may have a qualifier to disambiguate the meaning; qualifiers are put inside parentheses. The only punctuation marks that will be used are commas, to separate the family name from the first name which comes in this order: family, first; and parentheses for qualifiers. The index terms comes without any initial articles, unless for the proper, personal company and place names, if they have any; these names will be written as they are, as the people know them. For acronyms the dots will be omitted. Prepositions will not be used in general unless it is crucial for the meaning and without them the meaning could change. In addition, the first letter of a single term and the first letter of the first word in a compound word will be capitalized, unless it is a name where all first letters will be capitalized; capitalization will not affect the search. Last but not least, no inverted terms are used. Finally, all the index terms are post coordinated which will give the users more freedom and options when searching. We will always do our best to maintain consistency according to these guidelines.

**Alphanumeric arrangement**

Some of the related headings will not be collocated in the alphabetical list for example, elections and parliamentary elections. But this draw back will be rectified by the search option provided; users are able to search for the index terms they need using the search box on the same page of the alphabetical arrangement, which is also linked to our controlled vocabulary subject list. Each page of the alphabetical arrangement will have an indicator like which letter and page number. Terms with initial articles will be alphabetized as they are; the initial articles will precede the term and will be considered in the alphabetization; these terms will be indexed as they are, as the people know them for example, The Four Cats (band name). Moreover, the terms will be alphabetized word by word. Furthermore, numbers will be treated as numbers and will not be alphabetized. Many company names in Lebanon are written in French and this implies the use of modified letters; these modified letters will be alphabetized as the basic letters according to the NISO guidelines. The order used follows the NISO alphabetization guidelines where space comes first followed by punctuation followed by symbols followed by numbers and last come the letters. Some headings have qualifier to disambiguate the concept or meaning; qualifier will be considered when alphabetizing. All the subject terms are post-coordinated. On the design page 4 @! means symbols and # means numbers and each letter will take the users to the terms that starts with the same letter they choose accordingly.

**Locators and links to documents**

The locators in the Daily Star index are hyperlinks that link directly to the list of documentary units that have the same subject that we click on. For example, from the displayed index, alphabetical arrangement, list of records and surrogate records if we click on a subject heading like pollution it will take us to a list of records that have pollution as a subject heading. All the subject headings in the displayed index, alphabetical arrangement, list of records and surrogate records are hyperlinked (blue color). In the record and in the list of records the important locators, that emphasize a main theme, will be bold. This list will display the records in the regular view explained below in the surrogate display section; users are able to see all the bibliographic information in addition to the subjects and formats.

**Search interface**

The Home page has all the essential elements like name of index, small description of the index, overview of the contents, links to the displayed index, alphabetical arrangement, subject list, advanced search, site map, about us and the help and tips page.

The Daily Star index provides an organizing system with the classified index and categories, a navigation system where users can browse and a search system that allow user to search with a variety of options. From the Home page users will be able to browse and search as well. Top level categories are listed in the left sidebar and some sample headings. Users can choose to do a basic search or go to the advanced search page. Users can access the alphabetical arrangement, subject list (controlled vocabulary), browse by author or by country and access the search tips or help. Moreover, they can access the today issue and the archives as well, which go back for only two weeks; users are able to view the online version as it is, they can browse and read the past issue but they cannot search them yet. We give this option because of the embargo time the documentary units need to be indexed. Users can use the Boolean (and, or, not), Proximity (near) operators and phrase searching using the quotations. In the advanced search users will be able to perform field searching and they can add up to ten search boxes. In the Metadata we follow the Dublin Core but we had to make some changes for example we use author instead of creator and place instead of coverage because it is clearer to the users. They can access the tag cloud and the subject list (controlled vocabulary). Users have the option to limit their search before doing it; limiters are listed in the right sidebar (pre-search limiters). Similarly, users can filter the search after doing it as well (post-search limiters) the same limiters will be available on the result page plus some additional ones. Users are able to sort by relevance or by date.

**Surrogate displays**

In the result list the records will have the following displayed elements: title, author, source, date, issue, page, section, subjects, image showing the document type, add to folder, PDF or HTML sign (depends on the document), a very small abstract. Users can hover their mouse on “Expand” to view the full abstract. They also have the option to change the view of the records in the result list from brief to regular and full. The regular view is the one I described. The brief view will display only the bibliographic information of the document (title, author, date and section); users will be able to view more records on the same page. Not to mention that, the number of records can be controlled from 20 to 50 to 100, for all the views options. And the full view will give us a detailed view of the subjects, people, company, tags, place and the full abstract in addition to the elements from the regular view (full metadata).

As for each surrogate record, all the fields and metadata elements will be visible for the users. As mentioned above the Core Dublin elements are slightly edited so users don’t get confused. In addition, subjects will be differentiated from people as subject and place. Not to forget the tags. The full abstract is available plus all the options that users can benefit from after finding a record for example, save, print, cite, export, add to folder, and share (Facebook, Delicious, Twitter…). Users have the option to go back to the result list or jump from surrogate record to another directly.

On both the results page and surrogate page the search boxes will be available if the users need to edit or change the search.

**Metadata scheme**

Metadata will be stored in XML formats. With XML we can use the dynamic web; we can use HTML, CSS and JavaScript to edit the display and the layout without touching the metadata because it is stored in a separate XML file which we can linked to the HTML, CSS or JavaScript files. In addition, XML allows data sharing and gives structure to the data which is better and easier for retrieval. It is flexible, extensible and allows interoperability. We are using the Dublin Core schema because it is simple to use, you don’t need a specialist to create a record when using it. Furthermore, it is extensible which I think is very important because sometime we need to add new and specific elements for example in my case tags and people. One additional quality in DC is that all the elements are optional. Even though the DC has commonly known elements, we had to edit some of them so the users won’t be confused, for example author instead of creator and place instead of coverage. The 15 core elements provided by the DC with the addition of others we added are enough to cover and describe the documentary units published in the Daily Star.

**XML file**

<?xml version="1.0"?>

<metadata

xmlns="http://example.org/myapp/"

xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"

xsi:schemaLocation="http://example.org/myapp/ http://example.org/myapp/schema.xsd"

xmlns:dc="http://purl.org/dc/elements/1.1/">

<dc:documents>

<dc:document>

<dc:identifier>DS 3012/1</dc:identifier>

<dc:creator><b>Assaf, Nayla</b></dc:creator>

<dc:creator><b>Darrous, Sabine</b></dc:creator>

<dc:title>Both Murrs claim victory as Metn race goes down to the wire.

</dc:title>

<dc:date> 2002-03-03 </dc:date>

<dc:source>Daily Star</dc:source>

<dc:publisher>Kamel Mrowa</dc:publisher>

<dc:issue>3012</dc:issue>

<dc:pages>1</dc:pages>

<dc:language>English</dc:language>

<dc:type>text</dc:type>

<dc:format>online article</dc:format>

<dc:section><b>Politics</b></dc:section>

<dc:coverage><b>Lebanon</b></dc:coverage>

<dc:coverage><b>Metn</b></dc:coverage>

<dc:rights>public free access</dc:rights>

<dc:rights>no copyright law</dc:rights>

<dc:description>

The article talks about the parliamentary election that

took place in June 2002, for the Metn city. The article discusses

the process of the election and how it was affected by

politicians manipulations, bribery and corruption.

</dc:description>

<dc:full-text\_link>

<a href="http://www.dailystar.com.lb/News/Politics/Jun/03/Both-Murrs-claim-victory-as-Metn-race-goes-down-to-the-wire.ashx">Full-text</a>

</dc:full-text\_link>

<dc:subject>2002 elections</dc:subject>

<dc:subject><b>Bribery</b></dc:subject>

<dc:subject><b>Candidates</b></dc:subject>

<dc:subject><b>Corruption</b></dc:subject>

<dc:subject>Delegates</dc:subject>

<dc:subject><b>Elections</b></dc:subject>

<dc:subject><b>Freedom</b></dc:subject>

<dc:subject>Ministers</dc:subject>

<dc:subject><b>Parliamentary elections</b></dc:subject>

<dc:subject><b>Politics</b></dc:subject>

<dc:tags><b>Elections</b></dc:tags>

<dc:tags><b>Freedom</b></dc:tags>

<dc:tags><b>Justice</b></dc:tags>

<dc:tags><b>Murr election</b></dc:tags>

<dc:people><b>Murr, Elias</b></dc:people>

<dc:people><b>Murr, Gabriel</b></dc:people>

<dc:people><b>Murr, Michel</b></dc:people>

<dc:people><b>Murr, Mirna</b></dc:people>

</dc:document>

</dc:documents>

</metadata>

**References**

Law No. 75 of 1999 on the Protection of Literary and Artistic Property. *WIPO.* Retrieved from <http://www.wipo.int/wipolex/en/details.jsp?id=2786>

National Information Standards Organization. (1999). *Guidelines for alphabetical arrangement of letters and sorting of numerals and other symbols.* Bethesda, Maryland, USA: Wellisch, Hans.

Wacholder, N. (2013). *Classification and Categorization* Retrieved from <http://vizedhtmlcontent.next.ecollege.com/CurrentCourse/Class05-06-Classification%20and%20Categorization-LIS520,%20Fall%202013.ppt>

Wellisch, H. (1995). *Indexing from a to z*. (2nd ed.). New York: H. W. Wilson.