

Development of a Thesaurus as an Access Tool for the Jewish Networking Infrastructure Project

by

Rochelle Kedar

Jewish Agency for Israel and The Department of Information Science, Bar-Ilan University

I. Background: The Jewish Networking Infrastructure Project (www.jewishnet.org/sefer/)

The goal of the Jewish Networking Infrastructure Project (JNIP), which is an initiative of the Jewish Agency for Israel, is to establish basic infrastructures, which will enable the increased development of Jewish networking. The overall objectives are: (1) To develop a joint Jewish market for information, services and products leading to economies of scale in providing education, training, community development, collaborative decision making. (2) To lead to an increased level of interaction between different components of the Jewish People; reaching out and sustenance for Jewish Identity while developing unity and preserving diversity. To this purpose the project leaders have embarked on the organization of a broad coalition of Jewish umbrella organization and foundations, to engage their support for this project.

It is estimated that there are approximately 5,000 Web sites of Jewish interest, as well as several hundred mailing lists, numerous forums and chat, and several portals. These sites span a wide range of subjects: news, informational resources, education, distance learning, religious activities, community activities, political activism, commercial activity, leisure and travel, games, and more. Additionally, there exist many important Judaica databases that are outside of the reach of the Web.

The JNIP objectives will be facilitated through the integration of existing Jewish web sites and creating conditions for seamless access to a variety of media databases (sound, pictures, video), directories, indexes, catalogs and archives. This is being carried out through:

- (1) A comprehensive survey of Jewish Web sites.
- (2) Establishment of a Registry for Jewish networking resources.
- (3) Development of a controlled Jewish networking vocabulary.
- (4) Disseminating the use of an agreed upon Metadata Elements set for the description of existing resources.
- (5) Support for the operation of communities of knowledge, which will rate and annotate resources regarding their quality and pushing these annotated resources to Jewish content channels in general interest Portals.
- (6) Establishment of participatory mechanisms through which interested parties may participate in Working Groups and steer the development of the initiative.

II. Current Activities

1. **Completion of the Comprehensive Survey of Jewish Web Sites:** The JNIP team is currently compiling a list of addresses of most of the Jewish Web sites available, so as to encourage them to sign up in the Register of Jewish Networking resources. We estimate that some 5,000 such sites are available.
2. **Registry Development.** Will serve for: (a) Targeting the institutions, corporations and individuals that deal with Jewish networking. (b) Establishing a database with a description of the available resources. (c) To provide Web site owners with Metadata Tags they may choose to include describing their collections. (d) To disseminate a standardized form of description of Jewish Networking resources on the Internet (Dublin Core). (e) To establish a database enabling the participation of the target population in steering the project through Working Groups.

3. **Communities of Knowledge.** The work program envisions the operation of knowledge communities of experts rating and evaluating available documents. The best-rated resources will be channeled to Jewish Content Channels at general interest portals. The SharePoint package from Microsoft will be tested for this purpose. There are ongoing contacts with Sun to try also the iPlanet package for this purpose.

4. **Controlled vocabulary for Jewish Networking.** An initial version was presented in the International Congress for Jewish Studies in middle August. Feedback is being obtained from the managers of many Jewish Web sites. The vocabulary framework has been well received. The JNIP has acquired the right for the use a software package for management of multilingual controlled vocabularies from the INDIRE Institute in Firenze, Italy.

5. **OntoWeb** is a consortium of IT corporations, Telecoms and Universities for development of ontologies. The JNIP has joined OntoWeb, with the purpose to obtain technology, enabling agreements between Jewish archives and databases, so that the user can carry out seamless searches over different systems optimizing its information retrieval capacities. See: <http://www.ontoweb.org>

6. **Jewish Channels.** The JNIP is currently considering the immediate establishment of Jewish content channels. For this purpose: (a) the JNIP has received a proposal from MSN for establishing a Portal of Jewish Content Channels. (b) Negotiations with Google are being carried out for the establishment of a Jewish Google service. (c) Contacts have been established with the Open Directory Project, with the view of becoming one of its' Professional Content Providers.

7. **Seminar for Feedback and the Planning of the World Congress of Jewish Networking.** The JNIP Steering Committee decided to convene a seminar in December, 2001, with the owners and investors in some main Jewish sites so as to ascertain their feedback on the project's purposes and work plan. The P2P initiative and the Department of Education expressed their interest in collaborating in its organization. On the same occasion we will have a meeting of the Preparatory Committee for the World Congress of Jewish Networking.

III. The JNIP Thesaurus

In order to facilitate the efficient retrieval of materials, the Jewish Networking Infrastructure Project has embarked on the development of a thesaurus of indexing terms. The thesaurus will serve both the indexer and the searcher, and it will be available on the JNIP site. The major characteristics of this thesaurus will be:

Orientation: Jewish interest (Jews, Judaism, Jewish Life and Society, Israel)

Scope: all topics that would have any relation to Jews/Judaism etc.

Size: approximately 1200 descriptors (should not exceed 1500 descriptors)

Use: Post-coordination (Boolean searches)

Structure: Hierarchal, controlled vocabulary

Language: Multilingual (English-Hebrew, with the gradual addition of French, Spanish and Russian)

Access: through the Jewish Networking Infrastructure Project Gateway

The intended audience of the thesaurus, as the Gateway itself, is anyone from age 6 to 120, from grade school pupil to researcher, from housewives to professionals, who may be searching for material specifically relating to Jews, Judaism, etc., in a wide context. For example, specific topics will be indexed can include Jewish distance education, Jewish cuisine, Jewish community organization, institutions and services, Jews in Science and Technology, Jews in the Arts and Jewish Art, ... i.e., any topic that is related to Jews and/or Judaism.

The descriptors, in noun form, will be very clear and defined, with scope notes added as necessary. The indexer and the searcher will achieve specificity by the coordination of

one or more descriptors. For example, if one were interested in German-Jewish *Ketubot* (marriage contracts) of the 19th century, they would add to the above search AND *Ketubot*. In the English section of the thesaurus, English terminology will be used (e.g., Bible rather than Tanach), except in cases where the Hebrew term is judged as universally known (e.g., Tefillin rather than Phylacteries). In all cases, the appropriate cross-references will be constructed. Geographic proper names will be included in a separate list. Personal proper names will be added to the thesaurus as necessary.

If we compare with the circumstances of other multi-lingual thesauri, for example the Unesco Thesaurus, the official languages of the United Nations necessitate its multi-lingual character. In the case of the JNIP thesaurus, the character of the Jewish Diaspora necessitated its' multi-lingual aspect. The largest concentrations of Jews are situated in English speaking countries (US, UK, Canada, Australia etc), Israel, South America and Europe. It was therefore decided that initially the thesaurus would be English-Hebrew, with the future addition of Spanish, French and Russian. This will allow across-language searching: a search using Hebrew terms of the thesaurus, will retrieve documents in that language, as well as documents indexed with equivalent terms in the other languages.

Descriptors for the thesaurus will be chosen using a combination of methods:

- Bottom-up method – Examining web sites and sets of representative documents (titles, abstracts, or full texts), the extraction of important concepts and terms
- Top-down method – consultation with subject experts, use of dictionaries, glossaries and existing thesauri

After the formulation of a single preliminary set of terms a decision will be made as to the preferred terms and cross-references will be constructed from the variants and synonyms. Afterwards, hierarchical and associative relationships among the preferred terms (descriptors) will be defined. Finally a draft thesaurus (beta-thesaurus) will be produced, tested and revised. The thesaurus will be continually updated in order to insure its relevance and future efficacy.

Currently, the JNIP has produced a preliminary list of subject areas and topics to be included in the thesaurus. This list has been circulated among interested parties for comment. The major areas (top of hierarchies) included are:

Information sources (types of reference materials)

Information media

Information technology

Organizations and institutions

Judaism

Judaic sources

Holidays

Streams in Judaism

Rites, practice and customs

Jewish identity

Religious establishment

Folklore

Interdenominational relations

Education

Educational institutions

Pedagogy

Continuing education

Society

- Social groups
- Communities
- Family and home
- Politics and activism
- Public administration and services
- International relations

Economics

- Business
- Commerce

Language and literature

Arts

Recreation and leisure

Science and technology

Geography (division by country)

History (division by periods and centuries)

Zionism

The State of Israel

This preliminary subject-area list was compiled after looking over several representative websites. We are anticipating the need to index materials from all topic areas that could be of Jewish interest. Some of the specific topics that have been encountered:

Jewish Nobel laureates (www.yahoodi.com/famous/nobel1.html)

Occurrence of Cancer in Ashkenazi Jews in the US
(www.jewishsf.com/bk970829/uscolon.htm)

Interfaith dialogue (www.jcrelations.net)

Jewish firearm ownership (www.jpfo.org)

Jewish identity in contemporary theater (www.jewishculture.org/theater/9904-xroad.htm)

Jewish educational resources (www.jajz-ed.org.il)

Cyber-synagogue (www.primenet.com/~rabbilly)

Cyber-rabbi (www.pjtc.net/rabbis.html)

Our intention is that the thesaurus will cover subject areas that will be representative of all Jewish/Judaic topics, past, present and future, as they all appear in Cyberspace. While the collection of terms for the thesaurus does not seem to present any unforeseen problems, the establishment of hierarchal relationships may be subject to philosophical debate. For example, should Zionism be at the head of a hierarchy or should it be a narrow term of History? Another problem that we may encounter is as to the size of the thesaurus. We are interested in producing a thesaurus that will be small enough to be "handy" and yet achieve a reasonable level of specificity of terms. These are the challenges that await us. The initial list has generated much interest - we hope that the final product will live up to its' potential.