Peppering knowledge sources with SALT: Boosting conceptual content for ontology generation

Deryle Lonsdale*, Yihong Ding, David W. Embley**, Alan Melby***
*Department of Linguistics and English Language
**Department of Computer Science
Brigham Young University
Provo, UT 84602
lonz@byu.edu, {ding,embley}@cs.byu.edu, akm@byu.edu

Abstract

This paper describes work done to explore the common ground between two different ongoing research projects: the standardization of lexical and terminological resources, and the use of conceptual ontologies for information extraction and data integration. Specifically, this paper explores improving the generation of extraction ontologies through use of a comprehensive terminology database that has been represented in a standardized format for easy tool-based implementation. We show how, via the successful integration of these two distinct efforts, it is possible to leverage large-scale terminological and conceptual information having relationship-rich semantic resources in order to reformulate, match, and merge retrieved information of interest to a user.

Introduction

This paper describes work done to explore the common ground between two different ongoing research projects: one dealing with the development of a standardization scheme for lexical and terminological resources (SALT)[1], and the other for the establishment of a comprehensive infrastructure for leveraging and manipulating ontologies of declarative information types and conceptual relations for information extraction and data integration (TIDIE)[2]. Specifically, this paper focuses on how to substantially improve the generation of extraction ontologies by including a comprehensive terminology database that has been represented in a standardized format for easy tool-based implementation.

[1] Standards-based Access service to multilingual Lexicons and Terminologies (SALT), funded under EU Fifth Framework IST/HLT 3.4.1, addressing issues of international cooperation, standards for coding and interchange of linguistic data, and the combining of technologies (see www.ttt.org/salt/index.html).

[2] Target-based Independent-of-Document Information Extraction (TIDIE), funded under NSF Information and Intelligent Systems grant IIS-0083127 (see www.deg.byu.edu).

Two stages are discussed. First, we discuss the conversion of (portions of) a large-scale human terminological resource into a format that can be used by a data-extraction ontology system. Subsequently, we discuss how an ontology generation system integrates this terminological information with similar resources, and then uses the composite knowledge base to analyze the content of input documents and generate novel ontological relationships based on the observed concepts and their relationships. We discuss pertinent aspects of the generation engine including its knowledge base, the concept selection and conflict resolution processes, the retrieval of salient relationships, the discovery of related constraints, and the nature of the output ontology. We then discuss performance of the system in a particular subdomain (energy abstracts), and evaluation of the system's output. Finally, we draw conclusions and mention possible future applications.

Lexical resources and data modeling

A significant research area in the field of natural language processing (NLP) involves encoding lexical information in a format useful for a wide variety of computational processing modalities, from information retrieval to machine translation. The features and information in these lexical resources are designed to be used primarily by computers, and hence are often less than transparent to humans uninvolved in this kind of research. Widely varying data formats have emerged for these resources during their development by various research groups throughout the world for different languages and levels of linguistic description.

On the other hand the multilingual documentation industry involves the use of a wide variety of lexical resources that are designed and encoded specifically for use by translators, technical writers, and editors. These terminology databases (also called termbases) use a wide variety of models, are often concept-oriented in nature, and have been developed at great cost and effort (Wright & Budin, 1997). Widely known termbases include Canada's Termium, the European Union's Eurodicautom, and an ongoing terminology centralization project in the Nordic countries. Hundreds of standards bodies all over the world have created domain-specific termbases, and thousands of companies world-wide maintain their own corporate termbase(s). In fact, one of these standardization organizations, ISO, has created a multitude of domain-specific termbases.

There is considerable overlap between lexicons and termbases. For example, the Open Lexicon Interchange Format[1] is an intermediate format enabling the interchange between MT system lexicons (Thurmair et al., 1999), whereas MARTIF is an example of a termbase interchange framework between human end-users (ISO 12200, 1999). Current trends such as increased automation in language-related tasks and the integration of disparate knowledge sources have created a need for interchange between these lexical resources. Recent projects have focused on the issues inherent in integrating these types of resources (Melby & Wright, 1999), producing a new kind of highly structured information type that integrates both NLP lexicon and termbase data.

Given the wide variety of information to be treated and the range of formats currently in use, the field of lexicon/termbase data integration and exchange requires a principled approach to the modeling of data. The SALT project has introduced a data modeling approach that addresses the problem of interchange among diverse collections of such data, including their ontological substructure[2]. There are several interesting aspects of the approach; SALT provides:

- modularity, by differentiating a core structure from data category specifications;
- coherence, through the use of a meta-model; and
- flexibility, through interoperable alternative representations.

The modular meta-model approach has been implemented in various settings, and ongoing development is refining the model as it is tested more widely with more types of data.

For its part, data extraction also requires knowledge representations incorporating information that is terminological and conceptual in nature. To be ideal for generating extraction ontologies, knowledge sources must also: (1) be of a general nature, (2) contain meaningful relationships, (3) already exist in machine-readable form, and (4) have a straightforward conversion into XML. The TIDIE project's past work on extraction-ontology generation has used input knowledge from the MikroKosmos (μK) ontology[3] and general-purpose auxiliary data-frame libraries (i.e. application-independent and application-specific data frames, which are declarative descriptions of textual strings, including regular-expression recognizers, context-keyword recognizers, and applicable operations). The work reported in this paper, however, integrates information of a novel type: a large-scale terminology database which has some ontological structure and which has been reformatted according to the SALT standard and subsequently converted into μK-compliant XML for use by the ontology generator. In the next section we sketch this conversion process.

Termbase conversion

The Eurodicautom terminology bank[4] consists of over a million concept entries covering a wide range of topics. Each entry is multilingual in character, containing equivalents in any of several languages. Each entry is also associated with a rich array of the following types of information: sources cited, entry or approval dates, and so forth. Entries may contain single-word terms (e.g. "generator") or multi-word expressions (e.g. "black humus"). One crucial aspect of the information provided with each term is its Lenoch subject-area code. Lenoch codes are widely adopted in the documentation sciences as a hierarchical representation for classifying terms (and by extension their related concepts).

For the purposes of this paper the vast majority of information is not ultimately used; rather, only the English term and its subject code end up being relevant. Figure 1 shows part of the entry for a sample term ("black humus'); it shows translation equivalents for the term in various languages (Danish, Italian, Spanish, and Swedish). Crucially for this project, the line that begins with %%CM lists three Lenoch codes that associate pertinent subject areas with the term: AG4 (representing the subclass AGRONOMY), CH6 (representing ANALYTICAL-CHEMISTRY), and GO6 (representing GEOMORPH-OLOGY).

```
%%CM AG4 CH6 GO6
%%DA
%%VE lavmosetørv
%%RF A.Klougart
%%EN
%%VE black humus
%%RF CILF,Dict.Agriculture,ACCT,1977
%%IT
%%VE humus nero
%%RF BTB
%%ES
%%VE humus negro
%%RF CILF,Dict.Agriculture,ACCT,1977
%%SV
%%VE sumpjord
%%RF Mats Olsson,SLU(1997)
```

Figure 1: Partial entry from the Eurodicautom termbase in native format.

[1] See www.olif.net.
[2] In progress as ISO 16642 (forthcoming).
[3] See crl.nmsu.edu/Research/Projects/mikro/index.html
[4] See europa.eu.int.

The Eurodicautom format is simply one instance of a termbase encoding scheme, and is not usable as is for the purposes of data extraction as pursued in this paper. As a result, the data had to first be manipulated in such a way as to be rendered usable by the ontology generator.

Conversion of the termbase data was therefore performed via the SALT-developed TBX termbase exchange framework. Several thousand terms were converted to the TBX format, a refinement of the aforementioned MARTIF format which is based on the XML markup language. From there the terms could be converted to the XML format required by the ontology engine. The end result was a TBX-mediated conversion from native Eurodicautom terms to the final XML-specified ontology. Figure 2 shows a few sample terms resulting from the conversion process. Significantly, Lenoch codes (which are not understood by the ontology generator) have been re-interpreted as typical hierarchical relations which are understood by the generator (i.e. IS-A and SUBCLASS relationships).

```
<RECORD>
  <CONCEPT>xenobiotic substances</CONCEPT>
  <SLOT>SUBCLASSES</SLOT>
  <FACET>VALUE/FACET>
  <FILLER>hazardous raw materials </FILLER>
  <UID>0</UID>
</RECORD>

<RECORD>
  <CONCEPT>physical nuisances</CONCEPT>
  <SLOT>SUBCLASSES</SLOT>
  <FACET>VALUE/FACET>
  <FILLER>ambient light</FILLER>
  <UID>0</UID>
</RECORD>

<RECORD>
  <CONCEPT>financial statistics</CONCEPT>
  <SLOT>IS-A</SLOT>
  <FACET>VALUE/FACET>
  <FILLER>economic statistics</FILLER>
  <UID>0</UID>
</RECORD>

<RECORD>
  <CONCEPT>income</CONCEPT>
  <SLOT>IS-A</SLOT>
  <FACET>VALUE/FACET>
  <FILLER>budget</FILLER>
  <UID>0</UID>
</RECORD>
```

Figure 2: Sample converted Eurodicautom data as input to the ontology generator.

Ontology generation

In this section we discuss how the ontology is used in a system to extract information from text, and then to compare that information with the ontology itself. A sketch of the overall system, and its process for generating ontologies, is given in Figure 3.

Knowledge sources

Four kinds of basic knowledge sources were used in this work:

- the Mikrokosmos (μK) ontology, which consists of some 5,000 hierarchically-arranged concepts with a relatively high degree of connectivity (on average 14 inter-concept links per node) and which, by design, is general in nature and features inheritance of properties (Mahesh & Nirenburg, 1995);
- a data frame library, which is a repository of regular-expression templates designed to match structured low-level lexical items (such as measurements, dates, currency expressions, and phone numbers) and which can provide information for conceptual matching via inheritance (Embley et al., 1999);
- lexicons, which include onomastica (for matching geographic, personal, and corporate names) as well as the synset repository of WordNet (Fellbaum, 1998); and
- training documents, which contain domain-specific textual content of interest to a user, and which are assumed (for this paper) to be written in English.

Methodology

Given the abovementioned knowledge sources, prepro-cessing of the system's inputs can proceed (see Figure 3).

First, an integrated repository of conceptual information is created by mapping lexicon content and data frame templates to nodes in the merged ontology. This creates a unified and interconnected framework for matching lexical content from the input documents.

Secondly, the collection of training documents must be processed to extract its pertinent information. This step will vary according to the type of text available; in this paper, we assume appropriately-encoded HTML documents. These documents are first parsed to isolate linguistic content from other types of marked-up features. Record-detection processing may also be invoked where multiple records appear concatenated together without otherwise discernible separability. Then the linguistic content is tokenized and regularized. Once lexical content and textual input have been preprocessed, a four-stage generation process is invoked. This includes: (1) concept selection, (2) relationship retrieval, (3) constraint discovery, and (4) refinement of the output ontology. We discuss each of these stages in turn.

Concept selection. This stage involves finding which subset of the ontology's concepts is of interest to a user. Concepts are selected via string matches between textual content and the ontological data. Three different selection heuristics are performed to select concepts:

- concept-name matching
- concept-value matching
- data-frame pattern matching

String matches are calculated straightforwardly, with two additional assumptions: (1) word synonyms are considered via the use of WordNet synonym sets, and (2) multi-word terms undergo word-level matches. Hence CAPITAL-CITY is considered a synonym of both *capital* and *city*.

Concept-name matching selects concepts according to matches from conceptual names in the ontology's concept inventory. For example, suppose the sentence "Afghanistan's capital is Kabul and its population is 17.7 million." is contained within some document. Concept-name matching this sentence against the ontology would match the word "capital" with the concepts CAPITAL-CITY and FINANCIAL-CAPITAL and the word "population" with the concept named POPULATION. On the other hand, concept-value matching for the same sentence would match the word "Afghanistan" with an ontologically-specified value (or instance) for the concept COUNTRY-NAME and would match the word "Kabul" with an ontological value for the concept CAPITAL-CITY. Data-frame pattern matching against the same sentence would return a match for "17.7 million", selecting both the concepts POPULATION and PRICE. Clearly, matches occasionally entail incorrect conceptual interpretations for a given context, as in the case above where PRICE was matched. Once concepts have been selected, conflict resolution is performed on the results.

Concept conflict resolution seeks to arrive at an internally consistent set of selected concepts. Two levels of resolution are attempted: document-level resolution, and knowledge-source resolution. Though each is specialized and important to appropriate concept matching, for the purposes of this paper we only describe general principles that apply to both approaches.

Resolution leverages properties of lexical occurrence, proximity and distribution of words and terms for concept selection. Recall that in the previous example, both concept-name matching and concept-value matching generate CAPITAL-CITY for the word "capital", but only concept-name matching generates FINANCIAL-CAPITAL for the word "financial". According to the first resolution strategy, CAPITAL-CITY is retained and the FINANCIAL-CAPITAL is discarded due to the localized confluence of CAPITAL-CITY. Another resolution strategy is to prefer longer matches. For example, matching the term "bronze medal" would result in the concept ALLOY (generated via concept-value matching for the first word) and SPORT-ARTIFACT (generated via concept-value matching for the whole term). Resolution would result in SPORT-ARTIFACT being retained and ALLOY rejected. The third resolution strategy targets occasions where competing concepts are generated via the different knowledge sources. In the example above, recall that the number "17.7 million" resulted in both POPULATION and PRICE. In such instances the ontology itself might be used to resolve the conflict via any subsumption relationships that exist between the concepts. In our example, however, no hierarchical relation exists between the two, so this solution is not available. On the other hand, given the close proximity of proposed matched concepts, POPULATION, is preferred over PRICE. Other default strategies may also be invoked; discussion of these is beyond the scope of this paper.

Relationship retrieval. Once concepts have been matched and associated conflicts resolved, schemas representing the relationships between these concepts must be generated. This is accomplished by leveraging the conceptual relationships from the knowledge sources themselves.

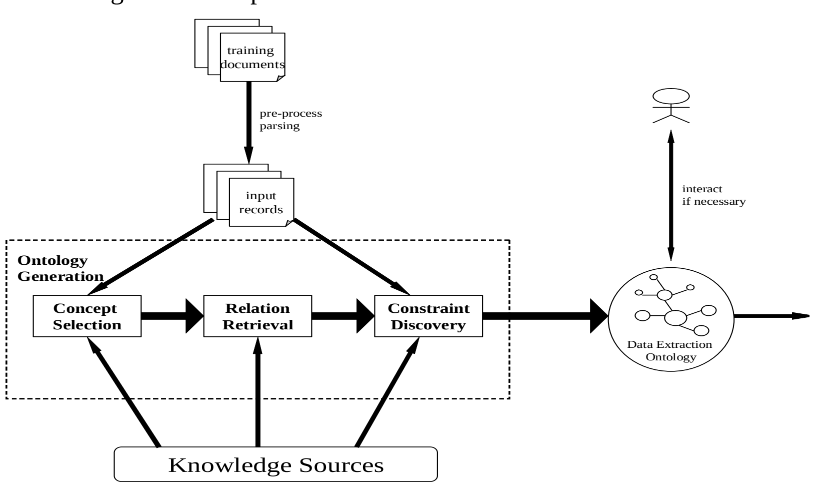

Figure 3: Architecture of the data-extraction ontology generation process

The ontology is structured as a directed graph whose nodes are concepts. All the concepts generated in the previous stage constitute a directed subgraph, either connected or unconnected, and the relationships among these concepts can be represented by paths among them. Theoretically, the best solution to finding the relationships among these concepts is to find all the paths in the subgraph; however, this is an NP-complete problem. Fortunately multiple paths, representing different relationships, rarely exist between any two nodes of interest. When they do, it is the shortest path that represents the most critical relationship between these two concepts. Note that the Onto-Search algorithm used in the µK project also adopts the shortest-path algorithm in relating two given concepts. Accordingly we adopt Dijkstra's algorithm, which is polynomial in complexity, to compute the most salient relationships between concepts. When such paths are too long (and hence their conceptual relationships consequently too weak) to be of interest, a distance threshold can be set to constrain relationship computation.

The next step is to construct schemas, or linked conceptual configurations, from the relationships posited in the previous step. First, a primary concept (i.e. the most important one) must be selected (or perhaps even posited, if it's deemed not to exist explicitly). Our technique selects the node with the highest degree of connectivity.

Constraint discovery. After concepts have been selected and their relationships established via the processing mentioned above, the system determines constraints on these relationships. For example, the cardinality of relationships between concepts can be constrained: people usually only have one birth date, but two parents, and potentially several phone numbers. The specification of cardinality constraints follows commonly adopted conventions (Embley, 1998), with a colon-delimited pair of symbols. For example, the constraint [0:1] on a given node associated with another in some relationship specifies that this given concept appears may in some but need not appear in all of the instances of that relationship (the 0: part), and that it appears no more than once in a given instance of that relationship (the :1 part).

Refining results. The output ontology may not be the final product that a user would want to deploy in a full-scale fashion for information extraction: hand-crafting of refinements and additions may need to be performed. Technically this can be done in a text editor since the ontology is a flat ASCII-encoded file; in reality, though, the end-user would need to be familiar with markup syntax and the specification of conceptual relations. In addition, the development of low-level matching predicates requires a familiarity with regular-expression writing. The use of ontology editors may help facilitate this final stage and bring ontology refinement within the realm of typical end-users. With rich enough knowledge sources and a good set of training documents, however, we believe that the generation of extraction ontologies can be fully automatic.

Results and evaluation

The system was run on various of U.S. Department of Energy abstracts[1]; an example abstract processed by the system is shown in Figure 4. The knowledge base used was the µK ontology, along with the Energy sub-hierarchy of Eurodicautom terms which contains almost 300 terms in this domain. A very small portion of the output ontology generated by the system according to the above process is given in Figure 5. In the system's output, all binary conceptual relationships in the ontology are represented one per output line, with cardinality constraints indicated as discussed above. Note that a new concept, energ2, was generated to represent the most relevant top-level concept.

A few remarks should be made about the generated ontology. Several dozen relationships are generated; of these, some are spurious and several are correct. For example, a relationship is appropriately posited between the concept CRUDE-OIL and the action PRODUCE; the role is Theme, meaning that one can PRODUCE CRUDE-OIL. This is a useful relationship; contrast this with the unhelpful relationship posited between GAS and GROW. Using traditional precision/recall measures is possible in assessing the number of inappropriate and appropriate matches. In our example, the high number of matches will result in relatively low precision figures. On the other hand, the system fares better when considered with respect to recall figures. This is not undesirable, however: it is easier for a human to refine the system's output by rejecting spurious relationships (i.e. deleting the false positives) than it is for a human to specify relationships that the system has missed.

This work relates loosely to NLP researchers' work in lexical chaining. The latter involves extracting and associating chains of word-based relationships from text, relating words and terms to resources like WordNet. The approach is particularly successful for applications like text categorization, automatic summarization, and topic detection and tracking (Green, 1999). Our main contribution here is in the grafting together of two disparate knowledge sources for similar tasks, and in generating a compatible set of ontological relationships which can then serve in a wide range of possible applications requiring hierarchically-structured information.

Several factors could contribute to an increase in the performance of the system. First of all, the µK ontology contains conceptual relations that favor a top-level view of the world and a particular set of assumptions about conceptual linkages. Arguably this is desirable in most cases, reflects the state of the art, and represents a good example of how generalized resources can function in various application areas. It follows that with a more

[1] The corpus is part of the ACL/DCI collection; see www.ldc.upenn.edu/Catalog/LDC93T1.html.

appropriate set of specified relations, though, our figures would improve; whether more appropriate resources even exist for a given application area is much less certain. Secondly, the Eurodicautom termbase information was also not designed to support this type of application, and accordingly problems result. As mentioned above, the Lenoch subject codes only allow for two types of hierarchical relationships: IS-A and its inverse, SUB-CLASSES. Our experience has shown that with a more varied set of conceptual relations, the generator is able to posit much more interesting relations. Note that the TBX format in fact contains several data types for encoding other types of relationships between concepts, so our work could be extended to other input formats. In fact, only the apparent paucity of publicly available ontologically-specified termbases at the current time stands in the way of further experimentation. It is expected, though, that a larger number of ontologically-annotated termbases will be increasingly made available in the near future. Finally, the nature of the texts themselves mitigates the success of the generation process to a certain extent. The texts chosen here are relatively free of data-rich information (e.g. low-level items such as dates, measurements, and names) that would otherwise allow for a richer set of concepts to be posited. Again, with more appropriate texts and a closer match with the input ontology, much better results can be achieved.

Still, the work of this paper is to show and explore the feasibility of even attempting to manipulate a termbase designed for human use into such a format that it can be used by a data-extraction system in order to discover and generate ontologically significant relationships. In this respect the system has succeeded, and we look forward to working with more focused and exhaustive termbase entries, more elaborate ontologies and richer input texts.

Conclusions

The importance of semantic information to natural language processing systems is increasing greatly, but acquisition of conceptual information is costly and difficult work. Fortunately, terminographers and lexicographers have often codified information that can be advantageous to work in semantic-based processing. With the successful integration of these two disparate areas, it is possible to leverage large-scale terminological and conceptual information with relationship-rich semantic resources in order to reformulate, match, and merge retrieved information of interest to a user. Possible future technology developed by this project can be embedded in personal agents, leveraged in customized search, filtering, and extraction tools, and used to provide individually tailored views of data via integration, organization, and summarization.

> The trend in supply and demand of fuel and the fuels for electric power generation, iron manufacturing and transportation were reviewed from the literature published in Japan and abroad in 1986. FY 1986 was a turning point in the supply and demand of energy and also a serious year for them because the world crude oil price dropped drastically and the exchange rate of yen rose rapidly since the end of 1985 in Japan as well. The fuel consumption for steam power generation in FY 1986 shows the negative growth for two successive years as much as 98.1%, or 65,730,000 kl in heavy oil equivalent, to that in the previous year. The total energy consumption in the iron and steel industry in 1986 was 586 trillion kcal (626 trillion kcal in the previous year). The total sales amount of fuel in 1986 was 184,040,000 kl showing a 1.5% increase from that in the previous year. The concept Best Mix was proposed as the ideal way in the energy industry. (21 figs, 2 tabs, 29 refs)

Figure 4: Sample Department of Energy Abstracts text, used as input to the ontology generation system.

Acknowledgements

The authors would like to thank Gerhard Budin for making the Eurodicautom data available for research purposes. Sergei Nirenberg was also very gracious in providing the μK ontology in various formats with its associated tools and documentation.

References

Embley, David W. (1998). *Object Database Development: Concept and Principles*. Addison-Wesley, Reading, MA.

Embley, D.W., Campbell, D.M., Jiang, Y.S., Liddle, S.W., Lonsdale, D.W., Ng, Y.-K., & Smith, R.D. (1999). Conceptual-model-based data extraction from multiple-record Web. *Data Knowledge and Engineering*, 31(3):227-251.

Fellbaum, Christiane (1998). *WordNet: An electronic lexical database*. Cambridge, MA: MIT Press.

Green, Stephen (1999). Lexical Semantics and Automatic Hypertext Construction. *ACM Computing Surveys* 31(4), December 1999.

ISO 12200:1999. Computer Applications in Terminology: Machine-readable Terminology Interchange Format (MARTIF) Negotiated Interchange. Geneva: International Organization for Standardization. (incl. amendment: ISO NP 12200:1999 Amendment 1, Extended MARTIF).

Mahesh, Kavi and Nirenburg, Sergei. (1995). A Situated Ontology for Practical NLP. In *Proc. Workshop on Basic Ontological Issues in Knowledge Sharing*, International Joint Conference on Artificial Intelligence (IJCAI-95), Aug. 19-20, 1995. Montreal, Canada.

Melby, Alan K. and Wright, Sue Ellen. (1999). Leveraging Terminological Data for Use in Conjunction with Lexicographical Resources. In *Proceedings of TKE '99: Terminology and Knowledge Engineering (23-27 August 1999, Innsbruck, Austria)*, Peter Sandrini (ed.), Vienna, pp. 541-570.

Budin, Gerhard & Melby, Alan K. (2000). "Accessibility of Multilingual Terminological Resources - Current Problems and Prospects for the Future". In *Proc. of the Second International Conference on Language Resources and Evaluation, Athens, Greece 31 May - 2 June 2000*, (Eds: E. Gavrilidou, G. Carayannis, S. Markantonatou, S, Piperidis, & G. Stainhaouer), Volume II, pp 837-844

Thurmair, Gregor; Ritzke, Johannes; and McCormick, Susan. (1999) The Open Lexicon Interchange Format: OLIF. TAMA '98—Terminology in Advanced Microcomputer Applications. In *Proceedings of the 4th TermNet Symposium: Tools for Multilingual Communication*. Vienna: TermNet. 237-262.

Wright, Sue Ellen, and Budin, Gerhard. (1997). *The Handbook of Terminology Management*, Vol. I. Amsterdam and Philadelphia: John Benjamins Publishing Company.

```
-- energy2 Information Ontology
energy2 [-> object];
energy2 [0:*] has Alloy [1:*];
energy2 [0:*] has Consumption [1:*];
energy2 [0:*] has CrudeOil [1:*];
energy2 [0:*] has ForProfitCorporation [1:*];
energy2 [0:*] has FossilRawMaterials [1:*];
energy2 [0:*] has Gas [1:*];
energy2 [0:*] has Increase [1:*];
energy2 [0:*] has LinseedOil [1:*];
energy2 [0:*] has MetallicSolidElement [1:*];
energy2 [0:*] has Ores [1:*];
energy2 [0:*] has Produce [1:*];
energy2 [0:*] has RawMaterials [1:*];
energy2 [0:*] has RawMaterialsSupply [1:*];
Alloy [0:*] MadeOf.SOLIDELEMENT.Subclasses MetallicSolidElement [0:*];
Alloy [0:*] IsA.METAL.StateOfMatter.SOLID.Subclasses CrudeOil [0:*];
Alloy [0:*] IsA.PHYSICALOBJECT.ThemeOf.PHYSICALEVENT.Subclasses Produce [0:*];
AmountAttribute [0:*] IsA.SCALARATTRIBUTE.MeasuredIn.MEASURINGUNIT
Consumption [0:*] IsA.FINANCIALEVENT.Agent Human [0:*];
ControlEvent [0:*] IsA.SOCIALEVENT.Agent Human [0:*];
ControlEvent [0:*] IsA.SOCIALEVENT.Location.PLACE.Subclasses Nation [0:*];
CountryName [0:*] NameOf Nation [0:*];
CountryName [0:*] IsA.REPRESENTATIONALOBJECT.OwnedBy Human [0:*];
CrudeOil [0:*] IsA.PHYSICALOBJECT.Location.PLACE.Subclasses Nation [0:*];
CrudeOil [0:*] IsA.PHYSICALOBJECT.OwnedBy Human [0:*];
CrudeOil [0:*] IsA.PHYSICALOBJECT.ThemeOf.GROW.Subclasses GrowAnimate [0:*];
CrudeOil [0:*] IsA.PHYSICALOBJECT.ThemeOf.PHYSICALEVENT.Subclasses Increase [0:*];
CrudeOil [0:*] IsA.PHYSICALOBJECT.ThemeOf.PHYSICALEVENT.Subclasses Combine [0:*];
CrudeOil [0:*] IsA.PHYSICALOBJECT.ThemeOf.PHYSICALEVENT.Subclasses Display [0:*];
CrudeOil [0:*] IsA.PHYSICALOBJECT.ThemeOf.PHYSICALEVENT.Subclasses Produce [0:*];
Custom [0:*] IsA.ABSTRACTOBJECT.ThemeOf.MENTALEVENT.Subclasses AddUp [0:*];
Display [0:*] IsA.PHYSICALEVENT.Theme.PHYSICALOBJECT.Subclasses Gas [0:*];
Display [0:*] IsA.PHYSICALEVENT.Theme.PHYSICALOBJECT.OwnedBy Human [0:*];
ForProfitCorporation [0:*] OwnedBy Human [0:*];
ForProfitCorporation [0:*] IsA.CORPORATION.HasNationality Nation [0:*];
Gas [0:*] IsA.PHYSICALOBJECT.Location.PLACE.Subclasses Nation [0:*];
Gas [0:*] IsA.PHYSICALOBJECT.ThemeOf.GROW.Subclasses GrowAnimate [0:*];
LinseedOil [0:*] IsA.PHYSICALOBJECT.ThemeOf.PHYSICALEVENT.Subclasses Increase [0:*];
```

Figure 5: Sample relations in the generated output ontology