Transferring ESRI Shapefile Data to LegendBurster and Working with LegendBurster Queries:

A Case Study using Mineral Deposits Data From Far East Russia, Alaska and the Canadian Cordillera[1]

GeoReference Online Limited

October, 2002

Abstract

This document explains how data can be transferred into a LegendBurster project from a conventional ESRI shapefile, and how that data may then be queried.

Examples illustrate the following special features of LegendBurster:

Tools for analyzing and reporting on the semantic structure of attributes associated with shapefiles;

The ability to work with hierarchically structured attribute values;

The ability to work with lists of attribute values;

The ability to display, as well as perfect matches, close and intermediate matches to queries.

---

[1] US Geological Survey Open File Report 98-136 (1998)

INTRODUCTION

In 1998 the US Geological Survey published[2] an important collection of maps on compact disk which related together the mineral deposits, metallogenic belts and structural terranes of the Russian Far East, Alaska and the Canadian Cordillera.

There are a number of aspects to this data set that make it a good candidate for transfer into a LegendBurster project - in particular its value to minerals exploration companies which need to understand the factors controlling the distribution of mineral deposits. Well-focused querying of the data base can provide important insights as to where exploration efforts should be directed.

Other reasons that make it an informative case study are

(a) the intrinsically hierarchical structure to some of the data it presents, and
(b) the "ranked-list" structure to other sections of its data.

Neither of these kinds of data are easy to query when stored in conventional relational database tables.

Storing the same data in semantic nets, however, allows for much more effective querying. In the LegendBurster environment, it also makes possible the display of different **degrees of matching** in the objects queried - an invaluable aid during data interpretation.

Most queries submitted to a conventional relational database return only the perfect matches to the query, making it difficult to represent, on the same map, the "near misses" that often exist, and are important to the final interpretation.

LEGENDBURSTER WORKFLOW

The flow of work during a LegendBurster project is shown graphically in Figure 1, and described in the paragraphs below:

[2] US Geological Survey Open File Report 98-136 (1998)

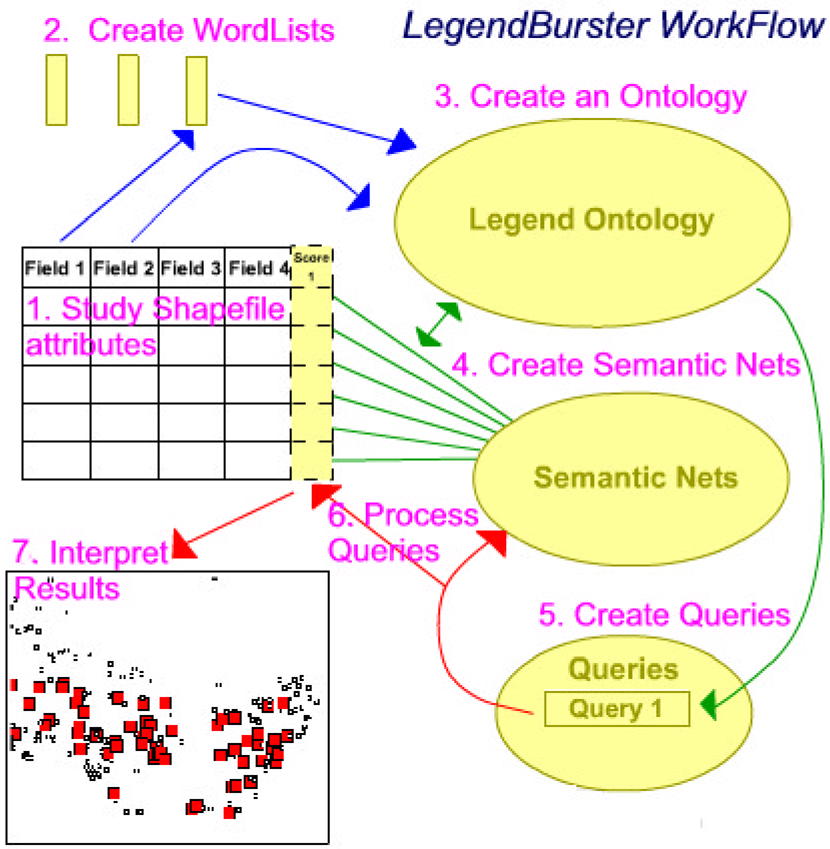


Figure 1: LegendBurster Workflow

(1) A detailed study is made of the structure of the data within each field of the table holding the shapefile attributes. Specific attention is paid:
a. to recognizing hierarchical structures in terminology used, and
b. to recognizing multi-valued attributes, and whether the multiple values are stored together in one field, or stored one-value-per-field.

(2) Using LegendBurster's tools for creating word lists from shapefile attribute tables, word lists are created for the map's attributes of interest. These are attributes which will usefully be used in queries, or which should appear together with query results, even though they may not be used as part of any query. Mineral deposit names, for example, very seldom form part of a query, but are essential in the display and reporting of query results.

(3) The word lists are used to create an ontology[3] for describing the features in the map. They may be supplemented by any related terminologies which may be imported into the project, either from another LegendBurster project, or from any suitably-formatted list of terms.

---

[3] "An ontology defines a common vocabulary for researchers who need to share information in a domain. It includes machine-interpretable definitions of basic concepts in the domain, and relations among them." (Noy and McGuiness (2001)). See Appendix 5 of this document for further explanation.

It is at this time that any hierarchical structure inherent to the data is given expression, which will in turn be utilized during querying of the data set.

(4) Each feature on the map (mapObject) is then described with its own semantic net, utilizing the terminology and semantic structures laid out in the ontology.

This is when multi-valued attributes are stored in a manner that allows them to be more precisely queried than is possible in most relational databases.

(5) Queries are then constructed using terminology and tools very similar to those used for the mapObject descriptions. For queries, this terminology includes words stating whether an attribute value should *always, usually, sometimes, rarely* or *never* be in the best matches to the query.

(6) Queries are then processed individually, or in batches.

Objects that exactly or closely match the query receive high scores, and objects without any matching attributes receive a low score. Intermediate matches receive intermediate scores. The scores for each query are stored in the project's shapefile for evaluation.

(7) Query results (scores) are then displayed in LegendBurster's Results Viewer, or in ArcView, ArcExplorer or ArcGIS 8. LegendBurster's Results Viewer has been optimally designed for investigating the "non-perfect" query scores - an essential task for the serious data analyst.

In the following pages, we will explain how the North Pacific Margin mineral deposits data referred to above was taken through these steps.

ANALYSING SHAPEFILE ATTRIBUTES, WORDLIST PREPARATION AND ONTOLOGY CREATION

The USGS Open-File Report 98-136 CD includes an ESRI shapefile called LODEDEP which holds the locational and attribute information for the deposits of the study. This shapefile is also part of a LegendBurster project called "TestLodeDep1" which may be downloaded from the LegendBurster section of the www.georeferenceonline.com web site. The TestLodeDep1 project includes all the semantic net data and queries which are described in this report.

The following two fields in the LODEDEP shapefile are of particular interest for this study :

Field Name	Description	Example Content
COMMODITIE	Deposit commodity(s), "listed in order of decreasing abundance or value".	"Ag, Au, Cu, Pb, W, Zn"
DEPOSIT_TY	"The type of lode deposit is an interpretation that was made by examining the summary of the deposit and then classifying the deposit using the deposit models previously described." - Quoted from text accompanying shapefile.)	"porphyry cu-mo and cu-au skarn"

The COMMODITIE field holds between one and eight commodities, which were quickly compiled into a list of unique values by LegendBurster's WordList tool (see Appendix 1). That list was then corrected for synonyms and invisible control characters present with some of the "cu" and "ag" values, and transferred to the Ontology Editor as permissible values for the attribute "Commodity". No hierarchical structuring was appropriate for this collection of values.

The DEPOSIT_TY field was also quickly compiled into a list of unique values using LegendBurster's WordList tool, and yielded 259 different values for deposit type, as shown in Appendix 2. It was apparent that structuring these values into a hierarchical classification system would enhance the value of the data set by allowing an examiner to query on either broad or narrow classes of deposit.

Two actions were therefore carried out:

(a) For demonstration purposes, the 259 deposit types were consolidated down to 43 classes, and entered manually into the project ontology. This was achieved primarily by dropping the commodity qualifiers, which are available anyway in the commodity field. Hierarchical structure was added to some areas of the new classification system by a drag-and-drop operation in LegendBurster's Ontology Editor - a procedure that took only a few minutes, but greatly enhanced the amenability of the data set to useful querying. The final classification system is shown in Appendix 3. (A serious mineral deposits researcher might generate a better, more complex hierarchical classification system from the original 259 deposit types in the shapefile. This was not done for reasons set out in (b) below.)

(b) Being aware that the USGS had published a comprehensive classification of mineral deposits of the world (Cox and Singer (1985)) which was available in LegendBurster format, the USGS classification was imported into the project's ontology as the preferred classification system (Appendix 4).

The project was now ready for the creation of semantic net descriptions of each mineral deposit.

DESCRIBING MAPOBJECTS WITH SEMANTIC NETS

The "Add Attribute to All …" procedure in LegendBurster was used to semi-automatically transfer the values in the COMMODITIE field of the shapefile into individual nodes of semantic net descriptions for each of the deposits in the database. (All LegendBurster procedures are fully explained in the system documentation integrated into LegendBurster, and accessible as context-sensitive "Help" screens.)

A new attribute called "RankNumeric" was then created in the project's ontology, and linked to the "Commodity" attribute, allowing each commodity to be given its rank, as determined from its position in the list of commodities. The "Add Attribute to All …" procedure was also used to semi-automatically add these values to the semantic nets.

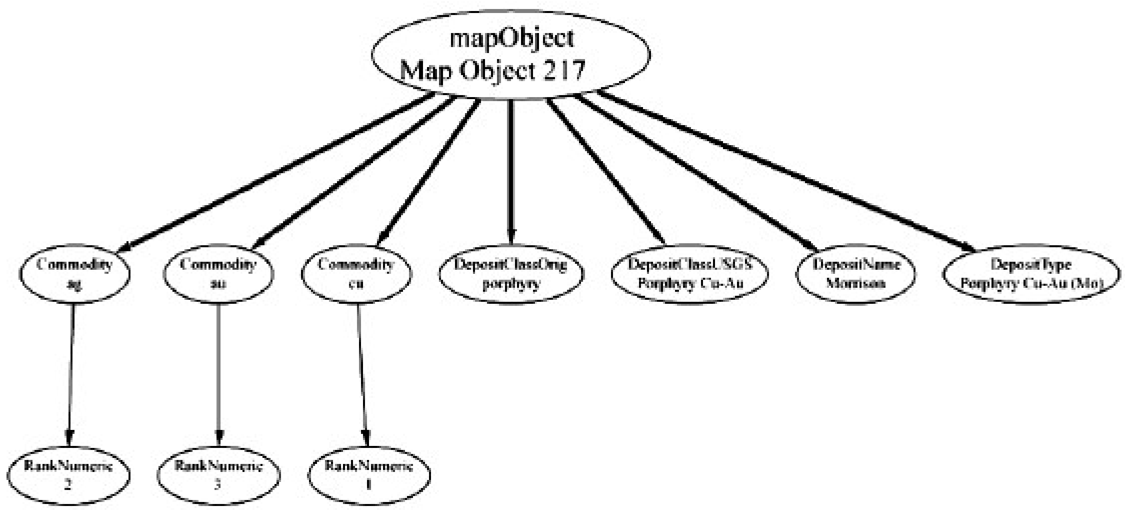
The same procedure was used to qualify ("describe") each deposit with two deposit-type attributes:

(a) "DepositClassOrig" - The simplified hierarchical classification developed for demonstration purposes.
(b) "DepositTypeUSGS" - The closest USGS deposit-type classification of the deposit, based on the author's interpretation of the DEPOSIT_TY value recorded in the shapefile for each deposit.

This allows the collection of mineral deposits to be queried using either the demonstration classification, or the internationally respected USGS classification.

As a next step, because the name of the deposit needs to be seen in the semantic net describing it, the deposit name (shapefile field: DEPOSIT_NA) was automatically transferred into all the nets using the "Direct ADD" procedure, and the same procedure was used to transfer the original deposit-type values into the semantic nets for reference purposes.

An example of one of the semantic nets at this stage of description is shown below:



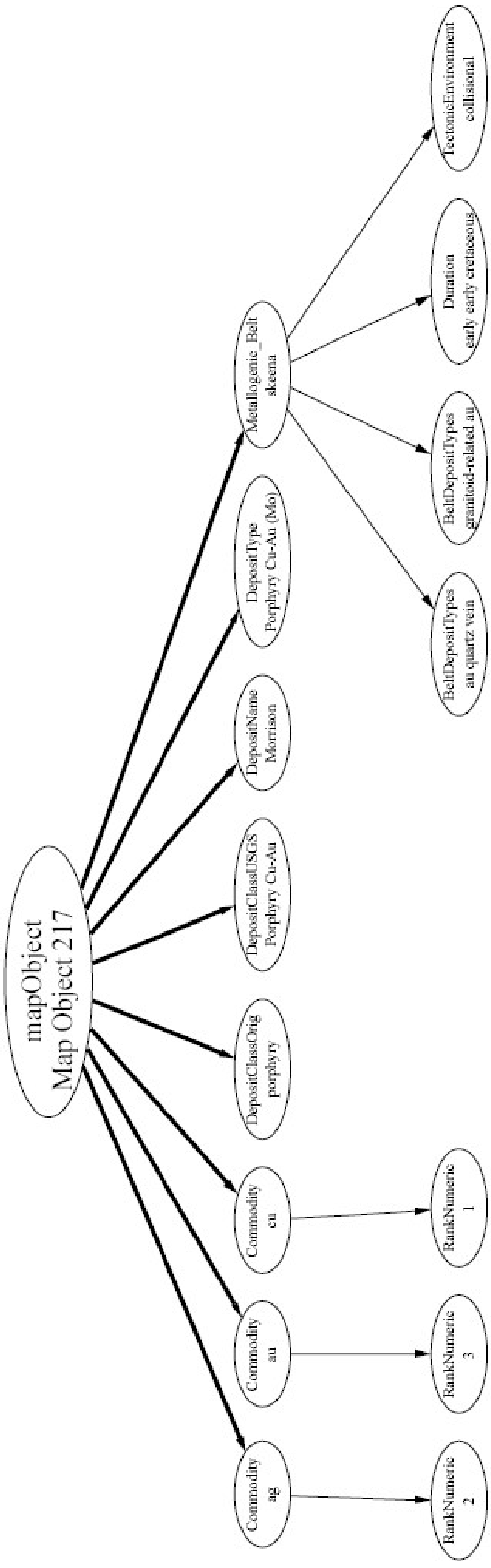
In 2001, the USGS published a table summarizing the ages of the metallogenic belts in the study region, the mineral deposit types each belt hosted, and the tectonic environment of each belt (USGS Open File Report 01- 261, Version 1).

Using a relational join based on the name of the metallogenic belt, this data was combined with the data discussed above, and, using the techniques discussed above, transferred to the LegendBurster project as attributes with the following names:

Metallogenic_Belt:	An attribute of the deposit (ie: of the mapObject in the above diagram).
TectonicEnvironment:	An attribute of the metallogenic belt: The tectonic environment in which the belt formed.
Duration:	An attribute of the metallogenic belt: The duration, in units of geological time, over which the belt was formed.
BeltDepositTypes:	An attribute of the metallogenic belt: The deposit types found within the belt.

This process was identical to the process which would be followed for a spatial join between points and any polygons within which they fall.

A diagram of the semantic net describing mapObject 217 after the additional attributes were added is shown below.



This additional data provided material with which to illustrate multi-level querying of semantic nets (Example 8 at the end of this document), as well as a data set illustrative of the challenges of dealing with periods of geological time which span more than one pre-defined time interval (to be discussed under separate cover).

QUERYING THE SEMANTIC NETS

LegendBurster has a unique approach to queries, whereby a query is considered to be a model of what is being sought. The model is specified by declaring how frequently it is expected to have certain relevant attribute values. The frequency values *always, usually, sometimes, rarely* and *never* are used for this purpose. The following two examples compare the natural language phrasing, SQL-type phrasing , and LegendBurster phrasing of the same query:

Natural Language:	Find all deposits with zinc as a commodity.
SQL-Type:	Find all deposits with “commodity” = “zinc”
LegendBurster:	The deposits I am looking for will always have zinc as a commodity

As illustrated in Example 1 on page __, the LegendBurster results are able to distinguish between deposits for which we have no information about copper, and deposits which are declared not to contain copper.

Natural Language:	Find all massive sulphide deposits in metallogenic belts which fall in continental terranes.
SQL-Type:	Find all deposits with “commodity” = “copper”
LegendBurster:	The deposits I am looking for will always be massive sulphide and will always come from a metallogenic belt which always falls in a continental terrane.

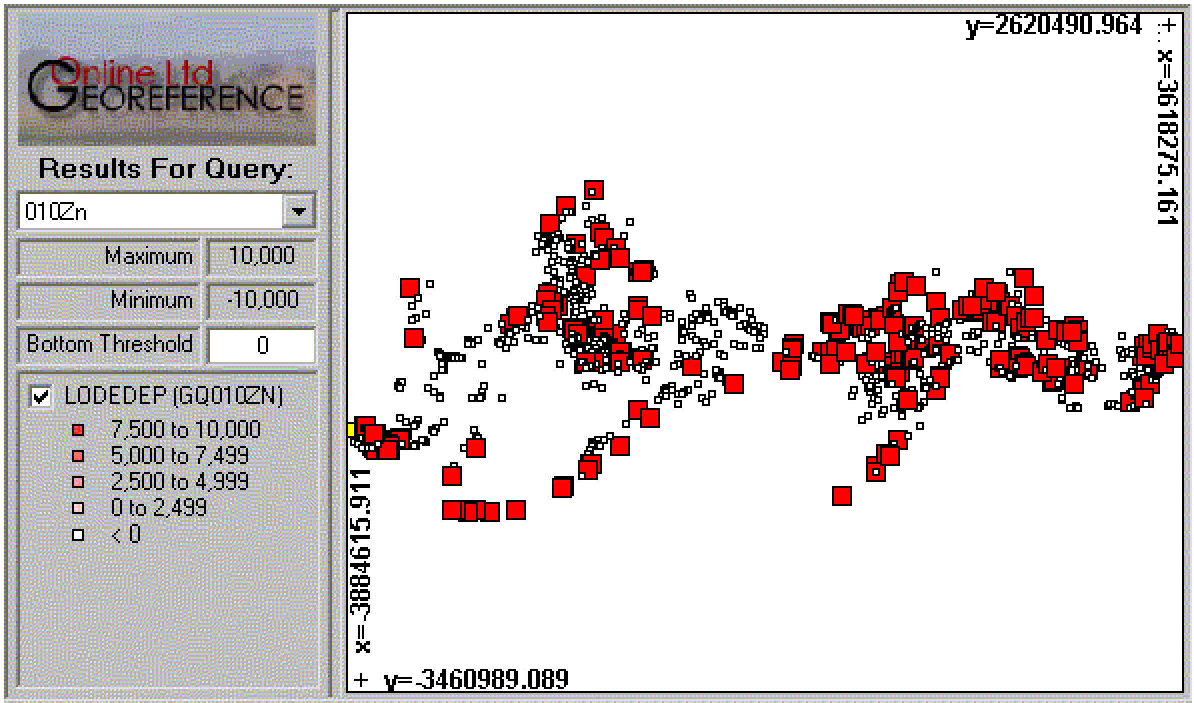
As illustrated in Example 8 on page __, the LegendBurster query identifies the deposits being sought, and also shows where massive sulphide deposits occur outside continental terranes, along with deposits which are not massive sulphide, but occur in continental terranes.

When working with lists of attribute values, the querying power of these frequency values can be enhanced in combination with the <others> value to represent the attribute values which have not been stated.

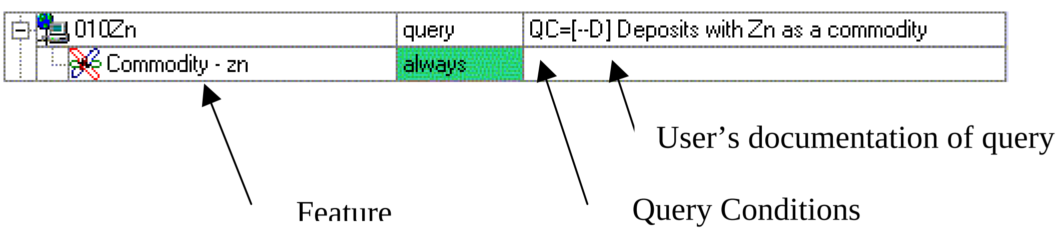
Thus, a search for all deposits containing commodities “Zn” and “Cu” would appear as shown in Example 5 below, while a search for all deposits containing only “Zn” and “Cu” would appear as shown in Example 6.

Query Example 1:

Show all deposits with zinc as a commodity:



LegendBurster Query:



The simplicity of this query results in only two scores being returned: 10000 for those deposits which have "zn" as a commodity, and 0 for those for which it is not stated whether "zn" is absent or present. Any deposit for which the absence of "zn" had been explicitly stated would have received a score of -10000.

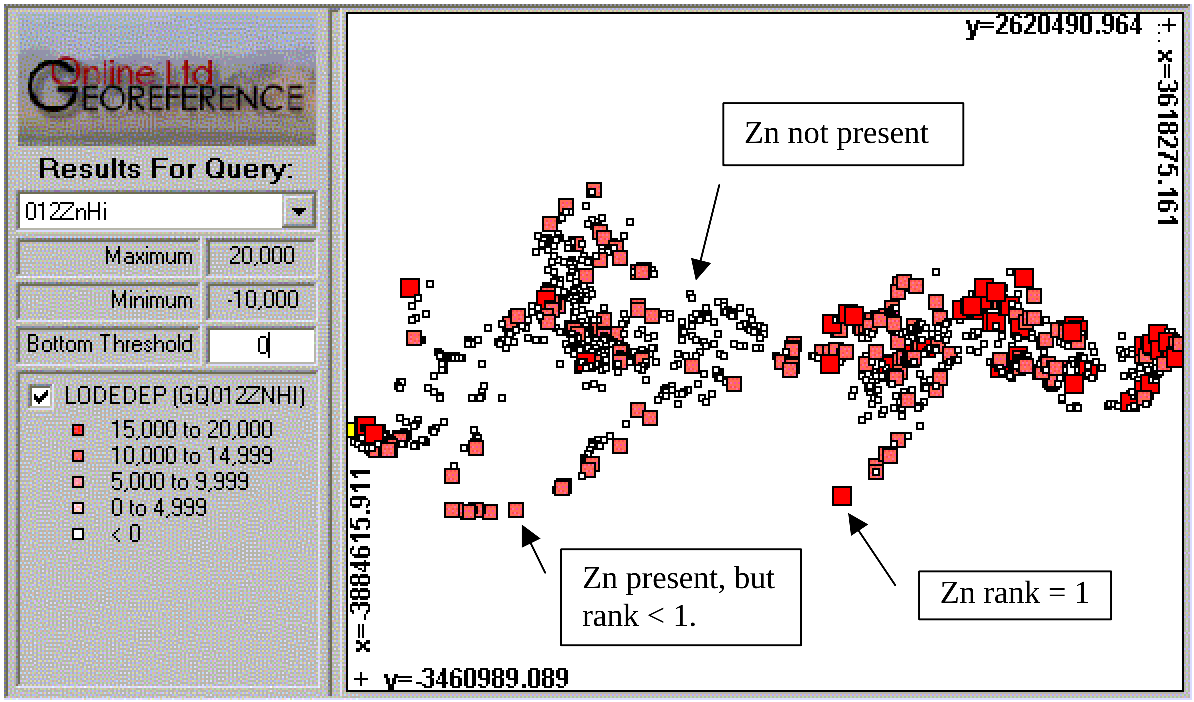
If the <silence implies absence> assumption had be selected, then all deposits without "zn" would have been given a score of –10000.

The Bottom Threshold has been set to 0 to enhance the difference between map symbols representing 10000 and symbols representing 0.

[Note that scores have not been normalized to the 0 – 100 range.]

Query Example 2:

Show all deposits with zinc as the top-ranked commodity.



LegendBurster Query:

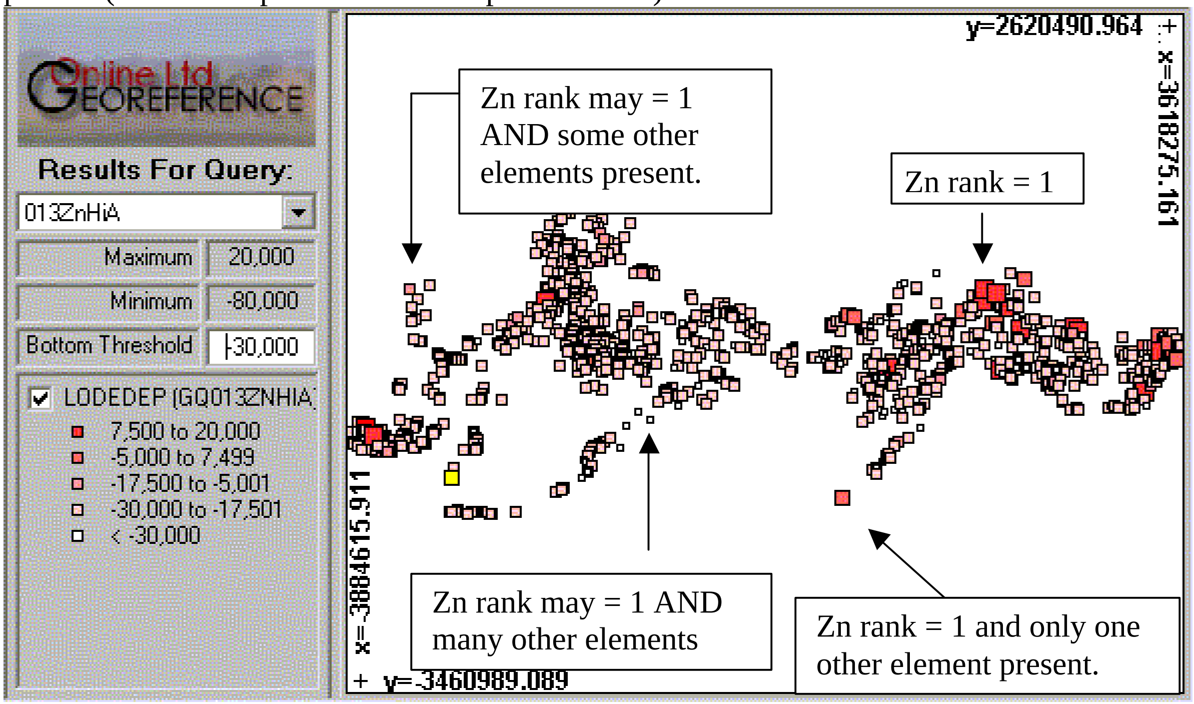
012ZnHi	query	QC=[--D] Deposits with Zn as the Rank 1 commodity
Commodity - zn	always	
RankNumeric - 1	always	

The maximum score possible is 20000, being 10000 for “zn” plus “10000” for a rank of 1. The estimated minimum is -10000 for instances in which zinc is declared to be absent.

The Bottom Threshold has been manually set to 0 to enhance the difference between map symbols representing 20000 and symbols representing 0. The system’s default setting for the Bottom Threshold was -4000 (being the -10000 minimum, plus 6000 - the bin range required for five bins between -10000 and +20000).

Query Example 3:

Show all deposits with zinc as the top-ranked commodity, without any other commodities being present ("silence implies absence" implementation).



LegendBurster Query:

013ZnHiA	query	QC=[-SD] Deposits with Zn as the Rank 1 commodity + Silence implies Absence
Commodity - zn	always	
RankNumeric - 1	always	

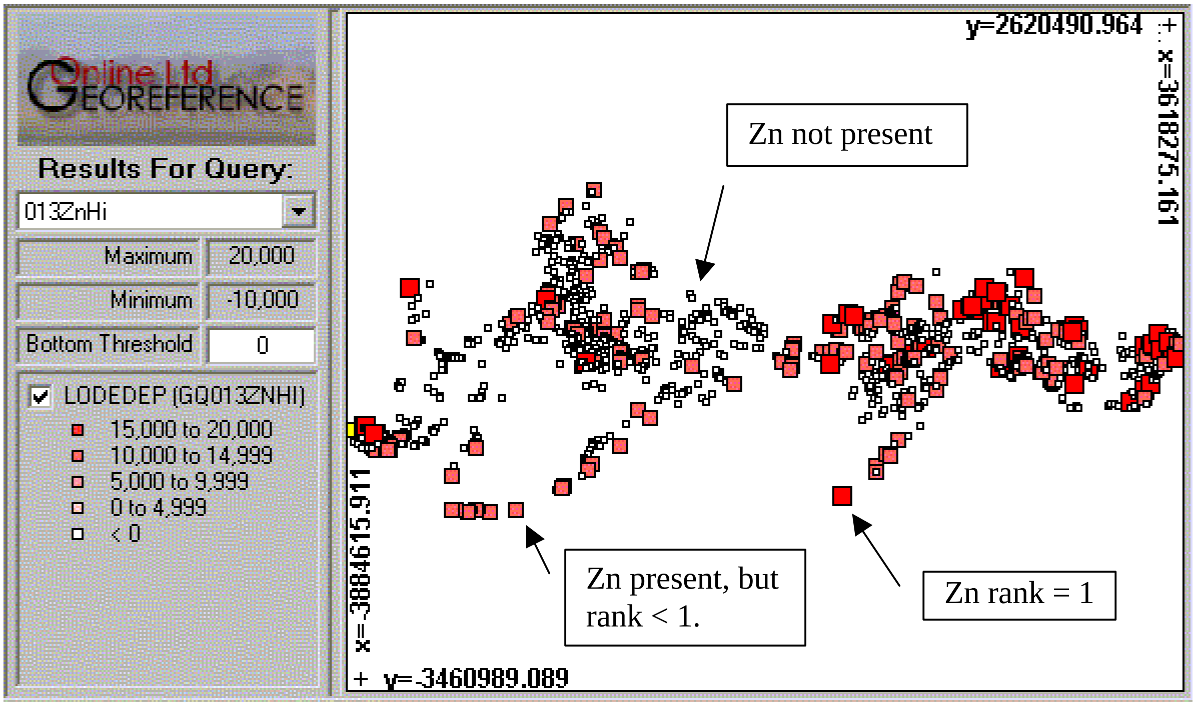
The maximum score possible is 20000, being 10000 for "zn" and "10000" for a rank of 1. The estimated minimum is -80000 for instances in which zinc is declared to be absent and seven other commodities are present in the instance (the maximum number of listed commodities in any one instance). The consideration of these "extra" commodities is a result of the "Silence implies Absence" Query Condition which has been set. The same scoring could have been achieved by adding the feature *Commodity <other values> "never"*. (Scoring methodology is explained in detail in the LegendBurster system documentation, and on the GeoReference Online Ltd web site.)

The Bottom Threshold has been manually set to -30000 to enhance the difference between scores in the range 20000 to -30000. The system's default setting for the Bottom Threshold was -60000 (being the -80000 minimum, plus 20000 - the bin range required for five bins between -80000 and +20000).

It is clear that the Query Conditions set for this query are not ideal, as it is returning a lot of information about additional commodities in mineral deposits in a context where this is not required. **The situation is clarified by adding to the query the information that "other commodities are possible (acceptable)", as illustrated in the next example**.

Query Example 4:

Show all deposits with zinc as the top-ranked commodity, not caring whether other commodities are present ("silence implies absence" implementation together with *<other values> "sometimes"*).



LegendBurster Query:

013ZnHi	query	QC=[-SD] Deposits with Zn as the Rank 1 commodity + Silence implies absence. Others allowed.
Commodity - <other values>	sometimes	
Commodity - zn	always	
RankNumeric - 1	always	

The maximum score possible is 20000, being 10000 for "zn" plus "10000" for a rank of 1.

The feature *Commodity <other values> "sometimes"* has been added to the query, which eliminates the need to accumulate penalties for commodities in any mineral deposit which are additional to zinc.

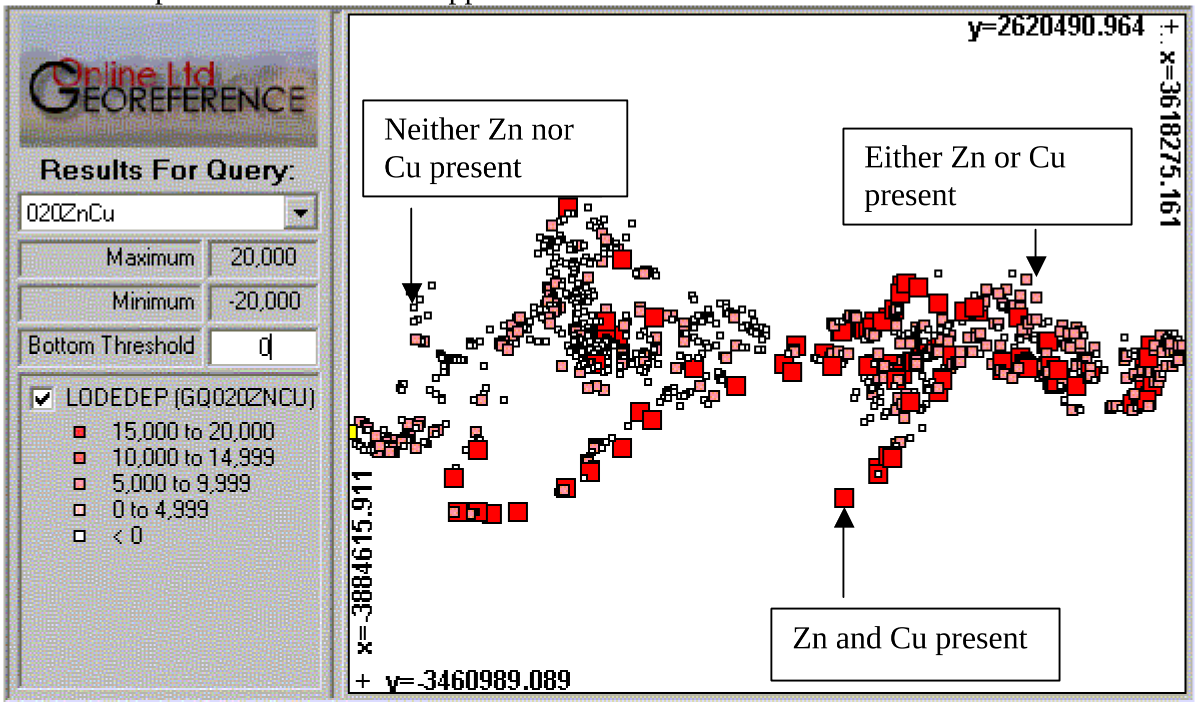
Hence the estimated minimum is -10000 for instances in which zinc is declared to be absent.

The Bottom Threshold has been manually set to 0 to enhance the difference between scores in the range 20000 to 0.

The results are essentially the same as those in Example 2. (However, if any attributes other than commodities are added to the query, "Silence implies Absence" will apply to them as illustrated in Example 3 (unless they, too, are "neutralized" with the *<other values> "sometimes"* feature).

Query Example 5:

Show all deposits with zinc and copper as commodities.



LegendBurster Query:

020ZnCu	query	QC=[--D] Deposits with Zn and Cu as commodities
Commodity - cu	always	
Commodity - zn	always	

The maximum score possible is 20000, being 10000 for “zn” plus “10000” for “cu”. The estimated minimum is -20000 for instances in which both zinc and copper are declared to be absent.

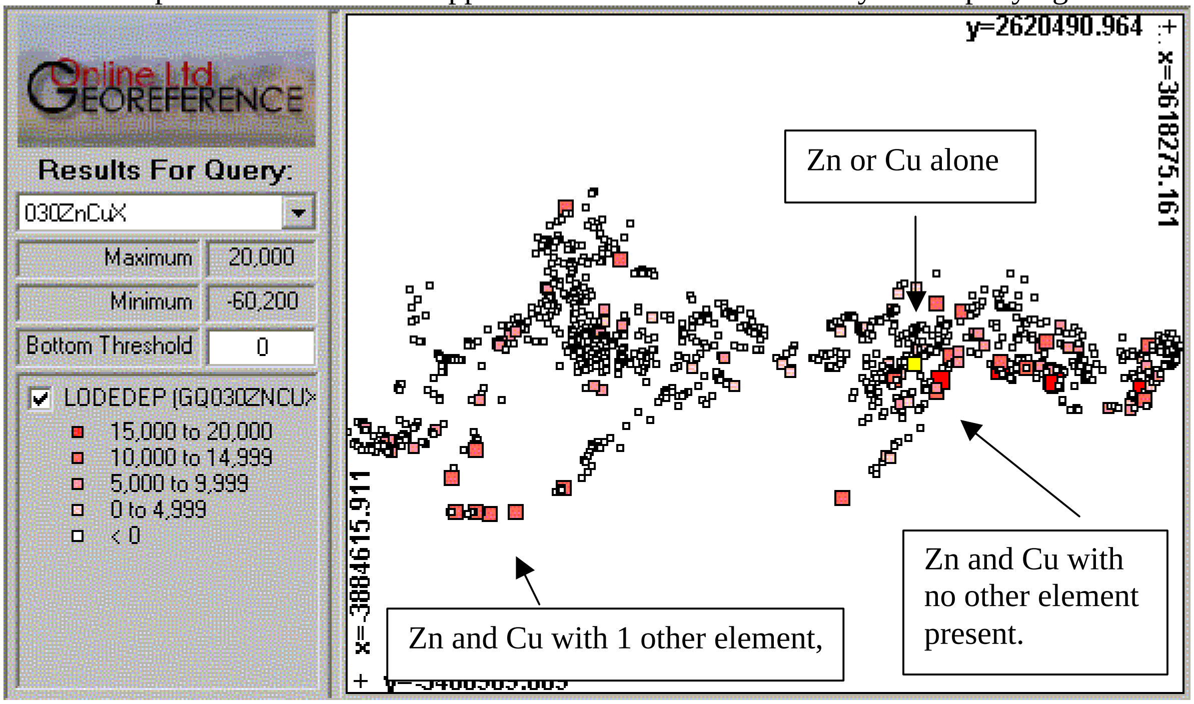
The Bottom Threshold has been manually set to 0 to enhance the difference between scores in the range 20000 to 0. The system’s default setting for the Bottom Threshold was -12000 (being the -20000 minimum, plus 8000 - the bin range required for five bins between -20000 and +20000).

The distribution of mineral deposits containing both copper and zinc is clear.

However, the distribution of copper and zinc existing independently of each other is also readily apparent in the map, and would not be shown as the result of any SQL command. A purpose-specific program would need to be written to generate a map similar to the one shown above.

Query Example 6:

Show all deposits with zinc and copper as commodities without any accompanying commodities.



LegendBurster Query:

030ZnCuX	query	QC=[--D] Deposits with Zn and Cu and no other commodities listed.
Commodity - <other values>	never	
Commodity - cu	always	
Commodity - zn	always	

The maximum score possible is 20000, being 10000 for “zn” plus “10000” for “cu”.
The actual minimum is -60200. This case is the lowest score actually recorded for the data set and is used because it is lower than the estimated minimum of -20000. The latter is estimated from summing penalties of -10000 each for deposits where both copper and zinc are declared to be absent.

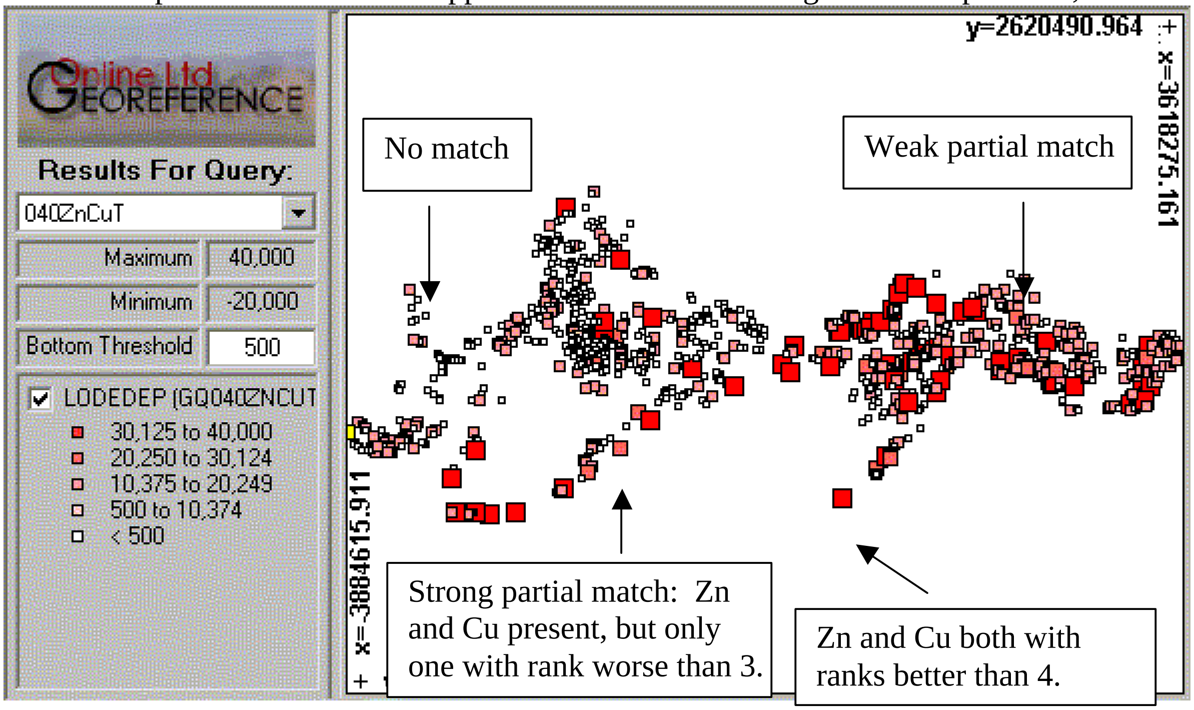
The Bottom Threshold has been manually set to 0 to enhance the difference between scores in the range 20000 to 0.

The distribution of mineral deposits containing both copper and zinc without any other commodities is clear.

However, the distribution of copper and zinc existing on their own is also readily apparent in the map, and would not be shown as the result of any SQL command. A purpose-specific program would need to be written to generate a map similar to the one shown above.

Query Example 7:

Show all deposits with zinc and copper as commodities ranking third in importance, or better.



LegendBurster Query:

040ZnCuT	query	QC=[--D] Deposits with Zn and Cu both above (ie: better than) rank 4.
Commodity - cu	always	
RankNumeric - <4	always	
Commodity - zn	always	
RankNumeric - <4	always	

The maximum score possible is 40000, being 10000 for "zn" plus "10000" for "cu", plus 10000 for each when they are ranked 3rd or better. The estimated minimum is -20000 for instances in which both zinc and copper are declared to be absent.

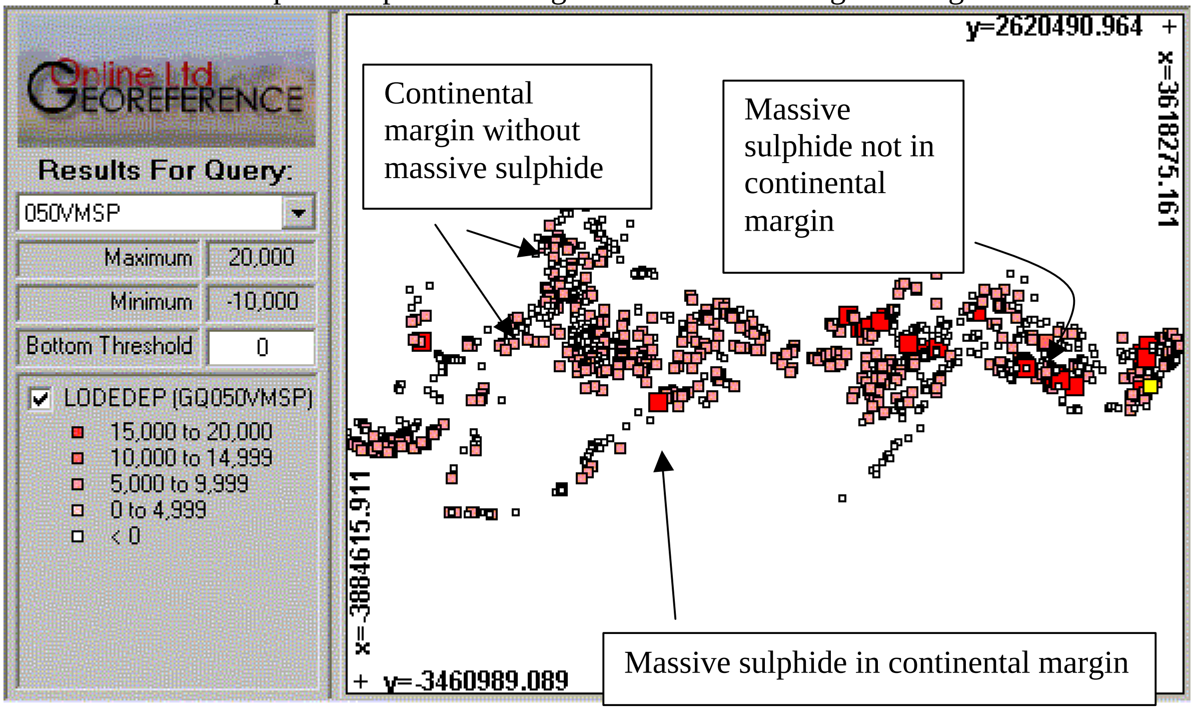
The Bottom Threshold has been manually set to 500 to enhance the difference between scores in the range 20000 to 0, in particular to ensure that scores of 40000 and 30000 fall in different bins.

Scores of 40000 are returned for deposits that satisfy all search criteria. Scores of 30000 are returned for deposits which have both required commodities, one of which is ranked below a level of 3.

The distribution of perfect matches to the query, and "close misses" is clear on the map, and would not easily be obtained from normal SQL queries.

Query Example 8:

Show all massive sulphide deposits existing in a continental margin setting.



LegendBurster Query:

050VMSP	query	QC=[--D] Masive sulphide deposits in belts classified as being in a continental margin setting.
DepositClassOrig - massive sulphide	always	
Metallogenic_Belt - <any value>	always	
TectonicEnvironment - continental margin	always	

This is an important example because it illustrates the essential role hierarchical classification systems can play in returning correct results to a query. As illustrated overleaf, a number of the perfect matches to the query are recognized only by referring to the classification hierarchies.

The maximum score possible is 20000, being 10000 for “massive sulphide” plus “10000” for “continental margin”. The estimated minimum is -10000 for instances in which massive sulphide is declared to be absent.

Since “metallogenic belt” may assume any value, it will never generate a penalty. And since estimated minimum scores do not need to take into account information below the first level in the query, no further penalties need to be taken into account.

The results display clearly the few massive sulphide deposits which occur in continental margin terranes (scores of 20000).

They also show massive sulphides in continental margins (scores of 10000), and, with different symbols, continental margin deposits which are not massive sulphides (scores of 9900; being 10000 for matching continental margin, and a penalty of -100 for not matching massive sulphide).

Using LegendBurster's comparison facility, it is possible to compare the query's features to those of any of the map's objects.

Shown below is a comparison of the current query with a matching mineral deposit.

050VMSP: Attribute	050VMSP's ...	050V...	: Attribute	's Value	: Pres...	Match Type
			Commodity	mo	present	
			Commodity	cu	present	
DepositClassOrig	massive sulphide	always	DepositClassOrig	massive sulphide kuroko	present	exactAKO
			DepositClassUSGS	Kuroko massive sulphide	present	
			DepositName	Red River	present	
			DepositType	Kuroko massive sulfide	present	
Metallogenic_Belt	<any value>	always	Metallogenic_Belt	tracy	present	exact
			BeltDepositTypes	Kuroko massive sulfide	present	
			Duration	late devonian	present	
TectonicEnvironment	continental margin	always	TectonicEnvironment	continental-margin arc	present	exactAKO

The comparison shows, in the "Match Type" column, that both matching attribute matches are "exactAKO" matches. This means that they have "a kind of (AKO)" relationships to the values they match in the query.

Without the ability to refer to the classification hierarchies, which is not present in most GIS systems, the correct results to this query would not have been achievable.

Appendix 1: "COMMODITIE" WordList

All Words in WordList: COMMODITIE

ID#	Word	Status	Related Word	Comment
0	ag	keep	-1	
1	ag	ignore	-1	Hidden control character
2	anhydrite	keep	-1	
3	as	keep	-1	
4	asbestos	keep	-1	
5	au	keep	-1	
6	au,ag	ignore	-1	
7	b	keep	-1	
8	ba	keep	-1	
9	barite	keep	-1	
10	bi	keep	-1	
11	caf2	synonym	21	
12	cd	keep	-1	
13	co	keep	-1	
14	cr	keep	-1	
15	cu	keep	-1	
16	cu	ignore	-1	Hidden control character
17	f	keep	-1	
18	fe	keep	-1	Is this only oxide iron? (See FeS and FeS2)
19	fes	keep	-1	See comment on Fe
20	fes2	keep	-1	See comment on Fe
21	fluorite	prefsynonym	21	
22	ge	keep	-1	
23	gems	keep	-1	
24	gypsum	keep	-1	
25	hg	keep	-1	
26	jade	keep	-1	
27	li	keep	-1	
28	magnesite	keep	-1	
29	mg	keep	-1	
30	mn	keep	-1	
31	mo	keep	-1	
32	nb	keep	-1	
33	ni	keep	-1	
34	p	keep	-1	
35	pb	keep	-1	
36	pd	keep	-1	
37	pge	keep	-1	

ID#	Word	Status	Related Word	Comment
38	phosphate	keep	-1	
39	pt	keep	-1	
40	ree	keep	-1	
41	s	keep	-1	Often listed with FeS and FES2. Clarify meaning.
42	salt	keep	-1	
43	sb	keep	-1	
44	se	keep	-1	
45	sn	keep	-1	
46	sr	keep	-1	
47	ta	keep	-1	
48	talc	keep	-1	
49	te	keep	-1	
50	th	keep	-1	
51	ti	keep	-1	
52	u	keep	-1	
53	v	keep	-1	
54	w	keep	-1	
55	w	keep	-1	
56	zn	keep	-1	
57	zr	keep	-1	

Appendix 2: “DEPOSIT_TY” WordList

Original WordList: DEPOSIT_TY

ID#	Word
0	ag epithermal vein
1	ag polymetallic vein
2	ag polymetallic vein and replacement
3	ag-au polymetallic vein
4	ag-co arsenide vein and fe-pb-cu-ag-au skarn
5	ag-pb-zn polymetallic vein
6	ag-pb-zn vein, polymetallic vein(?)
7	anorthosite apatite ti-p
8	as quartz vein
9	au epithermal vein
10	au polymetallic vein
11	au quartz and sb vein
12	au quartz vein
13	au quartz vein and au-sulfide disseminated
14	au quartz vein or polymetallic vein
15	au skarn
16	au sulfide disseminated
17	au-ag epithermal vein
18	au-ag epithermal vein
19	au-ag epithermal vein(?)
20	au-ag polymetallic vein
21	au-as polymetallic vein
22	au-co-as vein
23	au-pb-zn epithermal vein
24	au-pb-zn polymetallic vein
25	au-sb polymetallic vein
26	ba vein
27	ba vein and breccia
28	basaltic cu
29	basaltic cu and sediment-hosted cu
30	bedded barite
31	bedded barite, kuroko ba-zn-pb-cu massive sulfide
32	besshi massive sulfide
33	besshi massive sulfide(?)
34	boron skarn
35	carbonate-hosted hg
36	carbonate-hosted hg(?)
37	carbonate-hosted massive sulfide
38	carbonate-hosted zr (algoma type)
39	carbonatite-related ree
40	carbonatite-related ree (ta, nb)
41	clastic sediment-hosted hg
42	clastic sediment-hosted hg or hot-spring hg(?)
43	co skarn
44	co-as vein
45	co-bi-as vein
46	cu massive sulfide
47	cu skarn
48	cu vein
49	cu-ag quartz vein
50	cu-ag quartz vein(?)
51	cu-ag skarn
52	cu-au skarn
53	cu-fe skarn
54	cu-mo skarn
55	cu-pb-zn polymetallic vein
56	cu-pb-zn skarn
57	cu-zn skarn
58	cu-zn-pb-ba vein
59	cyprus massive sulfide
60	cyprus massive sulfide(?)
61	disseminated au-sulfide
62	epithermal vein
63	epithermal vein and volcanic-hosted sb vein
64	f vein
65	fe (au, cu, w, sn) skarn
66	fe (cu, pb, zn) skarn
67	fe skarn
68	fe-au skarn
70	felsic plutonic ree
71	felsic plutonic u
72	felsic plutonic u-ree
69	fe-pb-zn-sn skarn
73	fluorite greisen
74	gabbroic cu
75	gabbroic cu-ni
76	gabbroic ni-co
77	gabbroic ni-cu
78	gabbroic ni-cu(?)
79	granitoid-related au
80	granitoid-related au (ree)
81	granitoid-related au-ag (cu)
82	granitoid-related gold
83	hg quartz vein
84	hg-ag epithermal vein(?)
85	hornblendite peridotite cu-ni
86	hot-spring hg
87	ironstone
88	kennecott cu
89	kipushi cu-pb-zn
90	korean zn massive sulfide
91	kuroko cu-pb-zn massive sulfide
92	kuroko cu-zn massive sulfide
93	kuroko cu-zn-ag massive sulfide(?)
94	kuroko massive sulfide
95	kuroko massive sulfide(?)
96	kuroko massive sulfide(?) or polymetallic gold vein
97	kuroko pb-zn massive sulfide
98	kuroko zn-cu massive sulfide
99	kuroko zn-cu-pb massive sulfide
100	kuroko zn-pb-ag massive sulfide and zn skarn
101	kuroko zn-pb-cu massive sulfide
102	manto-replacement deposit (polymetallic pn-zn, au)
103	metamorphic ree(?)
104	metamorphosed sedimentary exhalative zn-pb(?)
105	mn-ag vein
106	mo greisen and vein
107	mo quartz vein
108	mo skarn
109	mo-cu skarn
110	paleoplacer u
111	pb polymetallic vein
112	pb-zn polymetallic vein
113	pb-zn skarn
114	pb-zn skarn and fluorite vein
115	pb-zn skarn and manto
116	pb-zn vein
117	pb-zn-(cu)-ag skarn
118	pb-zn-ag skarn and manto
119	pb-zn-ag skarn
120	pb-zn-ag vein
121	pb-zn-ag vein or skarn
122	pb-zn-au-ag vein
123	pb-zn-cu-ag skarn
124	podiform cr
125	podiform cr(?)
126	podiform cr-ni
127	polymetallic or epithermal vein
128	polymetallic sufide and au vein
129	polymetallic vein
130	polymetallic vein (metamorphosed)
131	polymetallic vein and au-ag breccia pipe or cu-au porphyry
132	polymetallic vein and kuroko massive sulfide
133	polymetallic vein and porphyry cu
134	polymetallic vein and porphyry cu(?)
135	polymetallic vein and porphyry cu-mo
136	polymetallic vein or cu-ag quartz vein
137	polymetallic vein or porphyry cu(?)
138	polymetallic vein or sb-au vein
139	polymetallic vein(?)
140	polymetallic vein(?), sn granite, porphyry mo
141	polymetallic vein, au quartz vein, sn skarn, cu-pb-zn skarn
142	polymetallic vein, au-quartz vein
143	polymetallic vein, pb-zn and possibly sn skarn
144	porphyry au
145	porphyry au-cu
146	porphyry co-mo

ID#	Word
147	porphyry cu
148	porphyry cu and (or) polymetallic vein
149	porphyry cu and cu skarn
150	porphyry cu and lesser polymetallic vein
151	porphyry cu and polymetallic vein
152	porphyry cu(?)
153	porphyry cu, polymetallic vein
154	porphyry cu-ag
155	porphyry cu-au
156	porphyry cu-au (mo)
157	porphyry cu-au and cu-au skarn
158	porphyry cu-mo
159	porphyry cu-mo and cu-au skarn
160	porphyry cu-mo and polymetallic vein
161	porphyry cu-mo and(or) polymetallic vein(?)
162	porphyry mo
163	porphyry mo-cu
164	porphyry mo-w, mo skarn
165	porphyry sn
166	porphyry sn or sn greisen
167	porphyry w-mo
168	porphyry-mo (w)
169	rhyolite-hosted sn
170	sb vein
171	sb-as vein
172	sb-au vein
173	sb-au vein (simple sb)
174	sb-au vein or clastic sediment-hosted sb-au
175	sb-au vein or polymetallic vein(?)
176	sb-au vein(?)
177	sb-au-hg vein
178	sb-hg vein
181	sedimentary exhalative zn-cu-pb
182	sedimentary exhalative zn-pb
183	sedimentary exhalative ba
184	sedimentary exhalative barite (pb-zn)
185	sedimentary exhalative ni-zn
186	sedimentary exhalative pb-zn
187	sedimentary exhalative pb-zn(?)
188	sedimentary exhalative zn-pb
189	sedimentary exhalative zn-pb-barite
190	sedimentary mn
191	sedimentary phosphorite
192	sedimentary zn-pb and (or) kuroko massive sulfide
179	sediment-hosted cu
180	sediment-hosted u
193	serpentine-hosted asbestos
194	silica-carbonate hg
195	sn (b) magnesian skarn
196	sn greisen
197	sn greisen and sn vein
198	sn greisen and vein
199	sn greisen with ta and nb
200	sn greisen(?)
201	sn greisen(?) and sn vein
202	sn polymetallic vein
203	sn polymetallic vein and greisen
204	sn polymetallic vein and sn silicate-sulfide vein
205	sn polymetallic vein(?)
206	sn quartz vein
207	sn quartz vein and greisen
208	sn quartz vein and sn greisen
209	sn silicate tourmaline, sn silicate-sulfide vein
210	sn silicate-sulfide and sn polymetallic vein
211	sn silicate-sulfide vein
212	sn silicate-sulfide vein and sn greisen
213	sn silicate-sulfide vein and sn polymetallic vein
214	sn skarn
215	sn-b skarn
216	sn-polymetallic vein, sn silicate-sulfide vein
217	sn-w greisen
218	sn-w polymetallic vein and greisen
219	sn-w skarn, sn greisen, carbonate -replacement sn(?)
220	southeast missouri ba-f
221	southeast missouri pb-zn
222	southeast missouri pb-zn and sediment- hosted cu
223	southeast missouri zn-pb-ag
224	stratabound fe
225	stratabound fe-mn
226	stratabound fe-p
227	stratabound gypsum
228	stratabound mg
229	stratabound w
230	strataform gypsum
231	strataform magnesite
232	stratbound gypsum
233	stratform salt
234	stratiform pb-zn
235	sulfur-sulfide
236	tin greisen and skarn
237	volcanic-hosted hg
238	volcanogenic fe
239	volcanogenic mn
240	w greisen(?)
241	w polymetallic vein
242	w skarn
243	w skarn and greisen
244	w vein and greisen
245	w vein, sn (w)-quartz vein
246	w-mo-sn vein and greisen
247	w-sn greisen
248	w-sn skarn
249	zn-ag polymetallic vein
250	zn-ag-au vein
251	zn-pb skarn
252	zn-pb skarn and manto
253	zn-pb-ag skarn and manto
254	zoned mafic-ultramafic
255	zoned mafic-ultramafic cr-pge
256	zoned mafic-ultramafic cu-au-pge
257	zoned mafic-ultramafic fe-ti
258	zoned mafic-ultramafic fe-v
259	zoned mafic-ultramafic ti

Appendix 3: Hierarchical Deposit-Type Classification System derived from "DEPOSIT_TY"

DepositClassOrig			DepositClassOrigCom
root			
	anorthosite		
	basaltic		
	bedded		
	breccia		
	carbonate-hosted		
	carbonatite-related		
	disseminated		
	felsic plutonic		including "granitoid-related"
	gabbroic		
	greisen		
	hot-spring		
	ironstone		
	kennecott		
	kipushi		
	manto		including "manto replacement"
	massive sulphide		
		massive sulphide besshi	
		massive sulphide cyprus	
		massive sulphide kuroko	
	metamorphic		
	podiform		
	porphyry		
	replacement		
	rhyolite-hosted		
	sediment-hosted		
		chemical sediment hosted	
		clastic sediment-hosted	
		paleoplacer	
		sedimentary exhalative	
		sedimentary phosphorite	
		stratabound	
		strataform	
	serpentine-hosted asbestos		
	silica-carbonate		
	skarn		
	southeast missouri		
	ultramafic		including " "hornblendite peridotite"
	vein		
		vein epithermal	
		vein quartz	
	volcanic hosted		
	volcanogenic		
	zoned mafic-ultramafic		

Appendix 4: Part of the (Hierarchical) USGS Classification of Mineral Deposit Types

DepositClassUSGS

- All_Deposits
  - Igneous
    - Extrusive rocks
      - Felsic-mafic extrusive rocks
        - Marine
          - Algoma Fe
          - Homestake Au
          - Kuroko massive sulphide
          - Low-sulfide Au-quartz vein
          - Volcanogenic Mn
          - Volcanogenic U
        - Subaerial
          - Deposits in older calcareous rocks
            - Carbonate-hosted Au-Ag
            - Fluorspar deposits
          - Deposits in older clastic sedimentary rocks
            - Almaden Hg
            - Hot-spring Hg
            - Silica-carbonate Hg
            - Simple Sb
          - Deposits mainly within volcanic rocks
            - Comstock epithermal vein
            - Creede epithermal vein
            - Epithermal Mn
            - Epithermal quartz-alunite Au
            - Hot-spring Au-Ag
            - Rhyolite-hosted Sn
            - Sado epithermal vein
            - Sn polymetallic veins
            - Volcanic-hosted magnetite
            - Volcanogenic U
      - Mafic extrusive rocks
        - Continental or rifted craton
          - Basaltic Cu
          - Sediment-hosted Cu
        - Marine, including ophiolite-related
          - Besshi massive sulfide
          - Blackbird Co-Cu
          - Cyprus massive sulfide
          - Komatiitic Ni-Cu
          - Volcanogenic Mn
    - Intrusive rocks
      - Alkaine and basic rocks
        - Alkaline intrusions in stable areas
          - Carbonatite
            - Alkaline complexes
            - Carbonatite model
            - Diamond pipes
      - Felsic intrusions
        - Mainly phanerocrystalline textures
          - Anorthosite intrusions

    - Anorthosite Ti
  - Granitic intrusions
    - Other wall rocks
      - Homestake Au
      - Low-sulfide Au-quartz vein
      - Sn greisen
      - Sn veins
      - W veins
    - Wallrocks are calcareous
      - Deposits far from contact
        - Carbonate-hosted Au
        - Polymetallic replacement
        - Replacement Mn
      - Deposits near contact
        - Carbonate-hosted asbestos
        - Cu skarn
        - Fe skarn
        - Porphyry Cu, skarn-related
        - Zn-Pb skarn
      - Replacement Sn
      - Sn skarn
      - W skarn
  - Pegmatitic
    - Be-Li pegmatites
    - Sn-Nb-Ta pegmatites
- Porphyroaphanitic intrusions present
  - High-silica granites and rhyolites
    - Climax Mo
    - Fluorspar deposits
  - Other felsic and mafic rocks including alkalic
    - Porphyry Cu
    - Wallrocks are calcareous
      - Deposits far from contact
        - Carbonate-hosted Au
        - Polymetallic replacement
        - Replacement Mn
      - Deposits near contact
        - Carbonate-hosted asbestos
        - Cu skarn
        - Fe skarn
        - Porphyry Cu, skarn-related
        - Zn-Pb skarn
      - Replacement Sn
      - Sn skarn
      - W skarn
    - Wallrocks are coeval volcanic rocks
      - In calcalkalic or alkalic rocks
        - Epithermal Mn
        - Porphyry Cu-Au
      - In granitic rocks in felsic volcanics
        - Porphyry Sn
        - Sn-polymetallic veins

Appendix 5: Definitions of "Ontology"

In Summary:

"An ontology defines a common vocabulary for researchers who need to share information in a domain. It includes machine-interpretable definitions of basic concepts in the domain, and relations among them."

From: "A Guide to Creating Your First Ontology" by Natalya F. Noy and Deborah L. McGuiness (Stanford University, Stanford, California) <View PDF file on WWW>

In More Detail:

"An ontology is a formal explicit description of concepts or classes in a domain of discourse, properties of each class describing various features and attributes of the classes, and restrictions on property values. An ontology together with a set of individual instances of classes constitutes a knowledge base. In reality, there is a fine line between where the ontology ends and the knowledge base begins."

Also from: "A Guide to Creating Your First Ontology" by Natalya F. Noy and Deborah L. McGuiness (Stanford University, Stanford, California) <View PDF file on WWW>

More Detail Still:

"An important and fundamental prerequisite to using a Representation and Reasoning System (RRS - described in Knowledge Bases) is to decide how a task domain is to be described. This requires us to decide what kind of things the domain consists of, and how they are to be related in order to express task domain problems. A major impediment to addressing this task is that there is no comprehensive theory of how to appropriately conceive and express task domains.

Despite this fundamental problem, the need for the following "commitments" is recognised:

- The world can be described in terms of entities (things) and relationships among entities. An ***ontology*** is a commitment to (or a formal declaration of) what exists in any particular task domain. This assumption that the world can be described in terms of things is the same that is made in natural language and logic. This isn't a strong assumption, as entities can be anything nameable, whether concrete or abstract. For example, people, colours, emotions, numbers and times can all be considered as entities. What is a "thing" is a property of an observer as much as it is a property of the world. Different observers, or even the same observer with different goals, may divide the world up in different ways.
- For each task or domain, one needs to identify specific entities and relations that can be used to express what is true about the world under consideration. **How one does so can profoundly affect one's ability to solve problems in that domain."**

Abridged from the final part of Section 1.3 of "Computational Intelligence: A Logical Approach" by Poole, Mackworth and Goebel (Oxford University Press (1998)). The sub-section is titled "Ontology and Conceptualisation".