

PATENTS ACT 1977

APPLICANT

Insightful Corporation

ISSUE

Whether patent application number GB
0518043.5 complies with section 1(2)

HEARING OFFICER

Mrs S E Chalmers

DECISION

Introduction

- 1 International patent application PCT/US2004/004099 was filed in the name of Insightful Corporation claiming priority from an earlier US patent application filed on 19 February 2003. The international application was published by WIPO as WO 2004/114163 on 29 December 2004 and entered the UK national phase as GB 0518043.5 on 12 February 2004.
- 2 The examiner has maintained throughout an objection that the invention claimed in this application is excluded from patentability as a computer program under s1(2) of the Patents Act 1977, which the applicant has not been able to overcome. The matter came before me for a decision on the papers. Although other objections were raised to the claims, this decision covers only the question of excluded matter, leaving these other questions to further processing of the application, if appropriate.

The application

- 3 The application is concerned with computerized systems for indexing and searching documents. The system is programmed to parse sentences to identify meaningful terms and their grammatical roles, to associate tags with those meaningful terms and to store representations of the tags. When a natural language query is received, the search engine parses the query and searches the indexed stored data set to locate data that contains similar terms used in similar grammatical roles and/or with similar entity tag types as indicated by the query. This syntactic approach is said to achieve more contextually accurate search results more frequently than using traditional search engines that work by generating results based on matching with words in the query.

4 The application as it currently stands is unamended although a proposed amended claim 73 was the subject of correspondence between the applicant and the examiner in an attempt to meet the patentability objection. The claims comprise a number of independent claims that relate to three distinct inventive concepts. These inventions are :

Invention I Methods, memory mediums and query engines for parsing, tagging and storing a tagged document as claimed in independent claims 1, 25, 49, 73, 86, 87, 90 and 91.

Claims 1 and 73 relate broadly to a computerized method for tagging a document by parsing sentences to identify meaningful terms and their grammatical roles, associating tags with those meaningful terms and storing representations of the tagged data. Claims 25 and 86 relate to a computer-readable memory medium containing instructions for tagging the document as specified in claims 1 and 73 respectively. Claim 49 relates to a syntactic query engine for carrying out the method of claim 1. Claim 90 relates to a computer-readable memory medium containing instructions for storing data tagged as specified in claims 1 and 73.

Invention II A data processing system as claimed in independent claim 89 comprising a memory containing structured tagged data.

Invention III Methods, memory mediums and query engines for searching documents and retrieving results as claimed in independent claims 93, 98, 103, 108, 113, 116 and 119.

Claims 93, 108 and 113 relate broadly to a computerized method for searching documents by parsing the query and searching an indexed stored data set to locate data that contains similar terms used in similar grammatical roles and/or with similar entity tag types as indicated by the query. Claims 98 and 116 relate to a computer-readable memory medium containing instructions for carrying out a search as specified in claims 93 and 113 respectively. Claims 103 and 119 relate to a query engine for carrying out the methods of claims 93 and 113 respectively.

Annex 1 to this decision recites the independent claims of Inventions I-III. Annex 2 to this decision recites the proposed amended claim 73.

The law and its interpretation

5 Section 1(2) of the Patents Act 1977 identifies certain types of subject matter for which patent protection is not available. The relevant provisions are:

1(2) It is hereby declared that the following (among other things) are not inventions for the purposes of this Act, that is to say, anything which consists of -

(a)

(b)

(c) a scheme, rule or method for performing a mental act, playing a game or doing business, or **a program for a computer;**

(d)

but the foregoing provision shall prevent anything from being treated as an invention for the purpose of this Act only to the extent that a patent or application for a patent relates to that thing as such.

- 6 The approved approach to determining whether matter is excluded is the four part test recently handed down by the Court of Appeal, in the *Aerotel and Macrossan*¹ case. The steps are as follows:
- a) properly construe the claim
 - b) identify the actual contribution (or, per paragraph 44 of the judgment, the alleged contribution will do at the application stage)
 - c) ask whether it falls solely within the excluded subject matter
 - d) check whether the actual or alleged contribution is actually technical in nature.
- 7 Paragraph 46 of the judgment adds that the fourth step may not be necessary because the third step may already have covered the point. This part of the test is in effect a longstop, to be invoked where the invention passes the first three steps.

Analysis

- 8 I shall first of all consider the present claims. If I find these to be unpatentable, I shall then go on to consider the proposed amended claim. At the outset, I should note that the agent has offered no argument in respect of the claims as filed.

Invention I

Construing the claims

- 9 I shall start by applying the first step of the test and I shall focus my attention on the method claims. I consider claims 1 and 73 relate to a computerized method for tagging a document by parsing sentences to identify meaningful terms and their grammatical roles, associating tags with those meaningful terms and storing representations of the tagged data.

Identifying the contribution made by the invention

- 10 As noted in the specification, typical search engines work by returning a list of sites that contain one or more of the submitted keywords with the sites with the most matches appearing at the top of the list. Relevant sites may therefore be missed if they don't return data matching with the submitted keyword. To try to improve the accuracy of search results, the application acknowledges that parsing has been used to identify the parts of speech of each term in a sentence to support the use of sentences as natural language queries and also to identify text sentences in a document that follow a particular part of speech pattern. The claimed invention not only parses the sentences in a document but also associates so-called "entity" tags with the parsed terms and stores

¹ Aerotel Ltd vs Telco Holdings Ltd & Macrossan's Patent Application [2007] RPC 7

representations of the tagged data for subsequent retrieval. These entity tags allow the parsed terms to be indexed not only according to their grammatical role but also by other attributes such as date, country or name to record the context in which the parsed term occurs. According to the specification, this enhanced way of tagging the data allows for significantly improved searching to locate more relevant search results. In my view therefore, the contribution made by the invention is a computerized method of tagging or indexing data to record both its grammatical role and its context and storing the data to provide for retrieval of search results.

Whether the contribution falls solely within excluded matter

- 11 It is clear to me on reading the specification that this invention is entirely concerned with setting up a computer to execute a series of operations in order to create what is in effect a database of tagged data. I agree that technical means, namely a computer, are used to implement the invention. However, that is not in itself enough to make the invention patentable. While the end result of implementing the invention is a new database, it is achieved by programming the computer to have the necessary functionality. Indeed it seems to me that the invention is closely analogous to *Macrossan* – the claims of which also include database structures – and that the contribution of the invention is similarly nothing more than a computer program up and running (see *Aerotel/Macrossan* at paragraph 73). Notwithstanding the form of words in which the invention is claimed, I therefore find the contribution falls squarely within the computer program exclusion.

Check whether the contribution is actually technical in nature

- 12 Having found in the third step that the contribution relates solely to excluded matters, it is not necessary, as I have explained above, for me to go on and consider whether it is technical in nature.

Invention II

- 13 The construction of claim 89 presents little difficulty. I consider it relates to a data processing system comprising a computer processor and memory containing tagged data stored as a table. I consider the monopoly sought is for a computer program and I can find no contribution in the claims that goes beyond the program. I therefore conclude that for the purposes of step (3), the invention is excluded as a computer program. Step (4) is therefore redundant.

Invention III

Construing the claims

- 14 As with my analysis of the first invention, I shall focus my attention on the method claims. I consider claims 93, 108 and 113 relate to a computerized method for searching documents by parsing the query and searching an indexed set of documents to locate data that contains similar terms to that of the query using latent semantic regression techniques.

Identifying the contribution made by the invention

- 15 As noted above in my assessment of the contribution of the first invention, typical search engines work by returning a list of sites that contain one or more of the submitted keywords. Relevant sites may therefore be missed if they don't return data matching with the submitted keyword. To try to improve the accuracy of search results, the application acknowledges that parsing has been used to identify the parts of speech of each term in a sentence to support the use of sentences as natural language queries and also to identify text sentences in a document that follow a particular part of speech pattern. The application also acknowledges that latent semantic regression is a known technique for searching documents to match a search query with a similar "match" at the sentence level in the located document. In the absence of a formal search, as far as I can determine, the contribution made by the invention is a computerized method of searching documents by parsing the query and searching an indexed set of documents to locate data that contains similar terms to that of the query using latent semantic regression techniques.

Whether the contribution falls solely within excluded matter

- 16 It is clear to me on reading the specification that this invention is entirely concerned with setting up a computer to execute a series of operations in order to search a corpus of documents and retrieve a set of results. I agree that technical means, namely a computer, are used to implement the invention. However, that is not in itself enough to make the invention patentable. While the end result of implementing the invention is said to achieve more contextually accurate search results more frequently than using traditional search engines, it is achieved by programming the computer to have the necessary functionality. Indeed it seems to me that the invention is closely analogous to *Macrossan* and that the contribution of the invention is similarly nothing more than a computer program up and running (see *Aerotel/Macrossan* at paragraph 73). Notwithstanding the form in which the invention is claimed, I therefore find the contribution falls squarely within the computer program exclusion.

Check whether the contribution is actually technical in nature

- 17 Having found in the third step that the contribution relates solely to excluded matters, it is not necessary, as I have explained above, for me to go on and consider whether it is technical in nature.

Proposed claim 73

- 18 Having found that all of the claims as filed fall within the computer program exclusion, I shall now consider the patentability of the proposed amended claim.

Construing the claim

- 19 The construction of the proposed claim presents little difficulty. I consider the claim broadly relates to a system for transforming a document comprising a natural language parser arranged to parse sentences to identify meaningful terms and their grammatical roles and to associate tags with those meaningful terms

and a data indexer arranged to store representations of the tagged data.

Identifying the contribution made by the invention

- 20 The agent asserts that the system as a whole is new, not because it is a computer program but because of the enhanced way in which data is stored which allows significantly improved searching through that data. As mentioned in my analysis of Invention I above, the use of parsing to try to improve the accuracy of search results is known. However, the claimed system not only parses the sentences in a document but also associates so-called “entity” tags with the parsed terms and stores representations of the tagged data for subsequent retrieval. In specific embodiments, the tagged data for each sentence may be stored in tabular form according to its grammatical role and entity type (eg country, organization). In my view therefore, the contribution made by the invention relates to a computerized method of tagging data to record its grammatical role and context and storing indexed representations of the tagged data.

Whether the contribution falls solely within excluded matter

- 21 The agent argues that the contribution provided by the way the tagged data is stored is “technical” as it results in an increased speed of processing by a search query engine and the enhanced retrieval of search results. However, it is clear to me, on reading the specification, that this invention is entirely concerned with setting up a computer to execute a series of operations in order to create what is in effect a database of tagged data. I agree that technical means, namely a computer, are used to implement the invention. However, that is not in itself enough to make the invention patentable. While the end result of implementing the invention is a new database, it is achieved by programming the computer to have the necessary functionality. The fact that the invention may result in an increased speed of processing seems to me to be an advantage rather than giving rise to a technical effect. Despite the agent’s assertion, I can find nothing in the specification that corresponds to the new “special exchange” of *Aerotel*. Indeed it seems to me that the invention is closely analogous to *Macrossan* – the claims of which also include database structures – and that the contribution of the invention is similarly nothing more than a computer program up and running (see *Aerotel/Macrossan* at paragraph 73). Notwithstanding the form of words in which the invention is claimed, I therefore find the contribution falls squarely within the computer program exclusion.

Check whether the contribution is actually technical in nature

- 22 Having found in the third step that the contribution relates solely to excluded matter, it is not necessary, as I have explained above, for me to go on and consider whether it is technical in nature.

Conclusion

- 23 I have found that that the invention as presently claimed relates to a computer program as such and is excluded from patentability under section 1(2)(c). I have also found that the proposed amended claim likewise relates to a computer

program as such. I have carefully read the specification but am unable to find anything that could form the basis of a patentable claim. I therefore refuse the application.

Appeal

- 24 Under the Practice Direction to Part 52 of the Civil Procedure Rules, any appeal must be lodged within 28 days.

MRS S E CHALMERS

Deputy Director acting for the Comptroller

Annex 1

Invention I

Claim 1. A method in a computer system for transforming a document into a canonical representation using entity tags, each entity tag having a type and an associated value, the document having at least one sentence, comprising:
receiving a designation of a plurality of entity tags; and
for each sentence,
parsing the sentence to generate a parse structure having a plurality of syntactic elements;
determining from the parse structure a set of syntactic elements that correspond to the designated entity tags; and
storing in an enhanced data representation data structure a representation of each association between a syntactic element of the determined set of syntactic elements and the type of the entity tag that corresponds to the syntactic element, the syntactic element representing the value of the corresponding entity tag such that the sentence is represented in the data structure by at least one entity tag.

Claim 25. A computer-readable memory medium containing instructions for controlling a computer processor to transform a document into a canonical representation using entity tags, each entity tag having a type and an associated value, the document having at least one sentence, by:
receiving a designation of a plurality of entity tags; and
for each sentence,
parsing the sentence to generate a parse structure having a plurality of syntactic elements;
determining from the parse structure a set of syntactic elements that correspond to the designated entity tags; and
storing in an enhanced data representation data structure a representation of each association between a syntactic element of the determined set of syntactic elements and the type of the entity tag that corresponds to the syntactic element, the syntactic element representing the value of the corresponding entity tag, such that the sentence is represented in the data structure by at least one entity tag.

Claim 49. A syntactic query engine for transforming a document into a canonical representation using entity tags, each entity tag having a type and an associated value, the document having at least one sentence, comprising:
parser that is structured to
receive a designation of a plurality of entity tags; and
decompose each sentence to generate a parse structure for the sentence having a plurality of syntactic elements;
determine from the structure of the parse structure a set of syntactic elements that correspond to the designated entity tags; and
store, in an enhanced data representation data structure, a representation of each association between a syntactic element of the determined set of syntactic elements and the corresponding entity tag type, such that the sentence is represented in the data structure by at least one entity tag.

Claim 73. A method in a computer system for transforming a document into a canonical representation using entity tags, each entity tag having a type and an associated value, the document having at least one sentence, each sentence having a plurality of terms, comprising:
receiving a designation of a plurality of entity tags and a designation of at least one grammatical role; and
for each sentence,
parsing the sentence to generate a parse structure having a plurality of syntactic elements;
determining a set of meaningful terms of the sentence from these syntactic elements;
determining from the structure of the parse structure and the syntactic

elements a grammatical role for each meaningful term;
determining which meaningful terms correspond to the designated entity tags
and which meaningful terms correspond to the designated grammatical role; and
storing in an enhanced data representation data structure a representation of
an association between the meaningful term that corresponds to the designated grammatical role
and an association between a meaningful term and the type of a corresponding designated entity
tag, the meaningful term associated with the entity tag type representing the value of the entity
tag, such that the sentence is represented by at least one entity tag and one meaningful term
having a grammatical role.

Claim 86. A computer-readable memory medium containing instructions for controlling a
computer processor to transform a document into a canonical representation using entity tags,
each entity tag having a type and an associated value, the document having at least one
sentence, each sentence having a plurality of terms, by:

receiving a designation of a plurality of entity tags and a designation of at least one
grammatical role; and

for each sentence,
parsing the sentence to generate a parse structure having a plurality of
syntactic elements;

determining a set of meaningful terms of the sentence from these syntactic
elements;

determining from the structure of the parse structure and the syntactic
elements a grammatical role for each meaningful term;

determining which meaningful terms correspond to the designated entity tags
and which meaningful terms correspond to the designated grammatical role; and

storing in an enhanced data representation data structure a representation of
an association between the meaningful term that corresponds to the designated grammatical role
and an association between a meaningful term and the type of a corresponding designated entity
tag, the meaningful term associated with the entity tag type representing the value of the entity
tag, such that the sentence is represented by at least one entity tag and one meaningful term
having a grammatical role.

Claim 87. A syntactic query engine for transforming a document into a canonical
representation using entity tags, each entity tag having a type and an associated value, the
document having at least one sentence, each sentence having a plurality of terms, comprising:

parser that is structured to
receive a designation of a plurality of entity tags and a designation of at least
one grammatical role;

decompose each sentence to generate a parse structure for the sentence
having a plurality of syntactic elements;

determine a set of meaningful terms of the sentence from the syntactic
elements; determine from the structure of the parse structure and the syntactic elements a role for
each meaningful term;

determine which terms correspond to the designated entity tags and which
meaningful terms correspond to the designated grammatical role; and

store, in an enhanced data representation data structure a representation of
an association between the meaningful term that corresponds to the designated grammatical role
and an association between a meaningful term and the type of a corresponding designated entity
tag, the meaningful term associated with the entity tag type representing the value of the entity
tag, such that the sentence is represented by at least one entity tag and one meaningful term
having a grammatical role.

Claim 90. A computer-readable memory medium containing instructions for controlling a
computer processor to store a normalized data structure representing a document of a data set,
the document having a plurality of sentences, comprising:

for each sentence,
determining a set of terms of the sentence that correspond to a designated set
of entity tags; and

storing sets of relationships between each determined term and its
corresponding entity tag type in the normalized data structure so as to represent the entire

sentence as entity tags.

Claim 91. A computer system for storing a normalized data structure representing a document of a data set, the document having a plurality of sentences, each sentence having a plurality of terms, comprising:

enhanced parsing mechanism that determines a set of terms of the sentence that correspond to a designated set of entity tags; and

storage mechanism structured to store sets of relationships between each determined term and its corresponding entity tag type in the normalized data structure so as to represent the entire sentence as entity tags.

Invention II

Claim 89. A data processing system comprising a computer processor and a memory, the memory containing structured data that stores a normalized representation of sentence data, the structured data being manipulated by the computer processor under the control of program code and stored in the memory as:

an entity table having a set of entity tag pairs, each pair having a term that is a value of a corresponding entity tag and an indication of an entity tag type of the corresponding entity tag.

Invention III

Claim 93. A method in a computer system for searching a corpus of documents, each document having a plurality of sentences, the corpus having an index of the plurality of sentences for the documents, comprising:

receiving an indication of a plurality of consecutive sentences;

parsing the indicated plurality of consecutive sentences to generate a plurality of search terms for searching the document corpus;

determining a plurality of result sentences in the corpus that correlate to the search terms using latent semantic regression techniques to determine the similarity of the search terms to the sentences in the corpus of documents; and

returning indications of the determined result sentences.

Claim 98. A computer-readable memory medium containing instructions for controlling a computer processor to search a corpus of documents, each document having a plurality of sentences, the corpus having an index of the plurality of sentences for the documents, by:

receiving an indication of a plurality of consecutive sentences;

parsing the indicated plurality of consecutive sentences to generate a plurality of search terms for searching the document corpus;

determining a plurality of result sentences in the corpus that correlate to the search terms using latent semantic regression techniques to determine the similarity of the search terms to the sentences in the corpus of documents; and

returning indications of the determined result sentences.

Claim 103. A query engine for searching a corpus of documents, each having a plurality of sentences, the corpus having an index of the plurality of sentences for the documents, comprising:

parser that is structured to

receive an indication of a plurality of consecutive sentences; and

decompose the indicated plurality of consecutive sentences to generate a plurality of search terms for searching the document corpus; and

postprocessor that is structured to

determine a plurality of result sentences in the corpus that correlate to the search terms using latent semantic regression techniques to determine the similarity of the search terms to the sentences in the corpus of documents; and

return indications of the determined result sentences.

Claim 108. A method in a networked computer environment for searching a corpus of documents, comprising:

receiving an indication of a plurality of consecutive sentences;
forwarding to a search engine the indicated plurality of consecutive sentences; and
receiving from the search engine indications of a plurality of result sentences from
the document corpus that correlate to the indicated plurality of consecutive sentences based upon
a latent semantic regression analysis used by the search engine to determine the similarity of
terms in the consecutive sentences to terms in the sentences of documents in the corpus.

Claim 113. A method in a computer system for searching a corpus of objects each object
having a plurality of units, the corpus having an index of the plurality of units for the objects,
comprising:

receiving an indication of a plurality of consecutive units;
decomposing the indicated plurality of consecutive units to generate a plurality of
search terms for searching the object corpus;
determining a plurality of result units in the corpus that correlate to the search terms
using latent semantic regression techniques to determine the similarity of the search terms to the
units in the corpus of objects ; and
returning indications of the determined result units.

Claim 116. A computer-readable memory medium containing instructions for controlling a
computer processor to search a corpus of objects each object having a plurality of units, the
corpus having an index of the plurality of units for the objects, by:

receiving an indication of a plurality of consecutive units;
decomposing the indicated plurality of consecutive units to generate a plurality of
search terms for searching the object corpus;
determining a plurality of result units in the corpus that correlate to the search terms
using latent semantic regression techniques to determine the similarity of the search terms to the
units in the corpus of objects; and
returning indications of the determined result units.

Claim 119. A search engine for searching a corpus of objects each having a plurality of
units, the corpus having an index of the plurality of units for the objects, comprising :

parser that is structured to
receive an indication of a plurality of consecutive units; and
decompose the indicated plurality of consecutive units to generate a plurality
of search terms for searching the object corpus and
postprocessor that is structured to
determine a plurality of result units in the corpus that correlate to the search
terms using latent semantic regression techniques to determine the similarity of the search terms
to the units in the corpus of objects ; and
return indications of the determined result units.

Annex 2

Proposed claim 73:

A system for transforming a document into a canonical representation using entity tags, each entity tag having a type and an associated value, the document having at least one sentence, each sentence having a plurality of terms, comprising:

- an input for receiving a designation of a plurality of entity tags; and a designation of at least one grammatical role;
- for each sentence,
 - a natural language parser arranged to:
 - parse the sentence to generate a parse structure having a plurality of syntactic elements;
 - determine a set of meaningful terms of the sentence from these syntactic elements;
 - determine from the structure of the parse structure and the syntactic elements a grammatical role for each meaningful term;
 - determine which meaningful terms correspond to the designated entity tags and which meaningful terms correspond to the designated grammatical role; and
 - a data indexer arranged to store in an enhanced data representation data structure a representation of each association between the meaningful term that corresponds to the designated grammatical role and an association between a meaningful term and the type of a corresponding designated entity tag, such that the sentence is represented by at least one entity tag and one meaningful term having a grammatical role.