



PATENTS ACT 1977

APPLICANT	Hitachi Vantara LLC
ISSUE	Whether patent application GB1714501.2 complies with the requirements of sections 1(1)(d) and 1(2)(c)
HEARING OFFICER	Mr B Micklewright

DECISION

Introduction

- 1 Patent application GB1714501.2, titled 'System for detecting data relationships based on sample data', was filed in the name of GB Gas Holdings Limited on 8 September 2017, making no claim to priority. Since filing, the registered applicant has changed twice, first to IO-Tahoe UK Limited on 9 January 2021 and then subsequently to Hitachi Vantara LLC on 7 December 2021. The application was published as GB2566931 A on 3 April 2019.
- 2 In the initial examination report, dated 28 June 2021, the examiner raised objections that the application was excluded from patentability as a computer program as such. Objections relating to plurality, novelty and inventive step were also raised, and the updating of the search was deferred until such time as the initial objections were overcome.
- 3 A response to this initial examination was received on 20 August 2021 which included amendments to both the claims and description. While the initial objections relating to novelty, inventive step and plurality were overcome by the amendments, several further rounds of examination and argued response followed on the issue of patentability. I note that for the last rounds of examination the examiner has deferred examination on all issues other than patentability, including completing the search for the application, until such time as the issue of patentability is resolved. The examiner included the offer of a hearing in their examination report of 8 February 2022, which was subsequently accepted by the applicant in their response of 28 March 2022. The matter came before me at a video hearing on 15 June 2022 with the applicant being represented by Dr John Addiss of Mewburn Ellis LLP. Also attending were observers Niall Deakin and Ketan Patel, and my assistant Charles Jarman
- 4 I am grateful for the skeleton arguments which were provided to me before the hearing. I confirm that I have taken account of these in reaching my decision. I have also reviewed the correspondence on file.

The Invention

- 5 The invention relates to a method of retrieving data from a data repository which includes a number of different data collections. When a number of different sets of data are imported into a common data repository, the structural information that would inform efficient querying of the information is often lost. Furthermore, there may be relationships between the various sets of data which would not have necessarily been previously defined.
- 6 An example provided within the application is that of a utility provider that might import e.g. a database of gas supply accounts and a database of boiler maintenance information from different parts of their organisation. The same customer may be present in both sets of data, with overlapping data items such as names, addresses, meter or boiler identifiers etc. appearing in both sets of data. It is not necessarily the case, however, that the same data will reside in the same tables/columns or that both datasets will be complete. Relationship information within a dataset may also be lost if only the raw data is copied into the new data repository.
- 7 The method involves using seed values, i.e. known data, which are used to identify sets of records from both first and second data collections in which the seed values can be found. These sets of records are then searched for common values, and if common values are found then a candidate relationship between the respective fields of the first and second data collections is outputted. This identified candidate relationship can then be used in the creation and execution of queries to retrieve data from the data repository, with the query including a join predicate that specifies an equivalence or other relationship between the first and second fields. Where a common value is not initially identified when searching the identified records, the search is iteratively expanded and repeated in one of the first or second record sets using one or more other values appearing in the record set as an additional seed value until either a common value is identified or a termination criterion is met.

The claims

- 8 Independent claim 1 states:
 1. A method of retrieving data from a data repository based on identifying relationships between given data collections of a plurality of data collections in the data repository, each data collection comprising a plurality of data records, each data record comprising one or more data fields storing one or more data values,

wherein the data repository is connected via a network to a user device and to a plurality of data sources, the data collections imported from the data sources via the network using a data import tool;

the method comprising performing a search process including:

receiving a first seed value and a second seed value;

identifying and retrieving a first set of records from the plurality of data collections based on the first seed value;

identifying and retrieving a second set of records from the plurality of data collections based on the second seed value;

searching for a common value across the retrieved first and second record sets, wherein the common value is a value which appears in a first field in a first record of the first record set and in a second field in a second record of the second record set, wherein the first record is from a first data collection and the second record is from a second data collection;

in response to not identifying the common value when searching using the retrieved records, iteratively expanding one of the first and second record sets using one or more other values appearing in the record set as additional seed values and repeating the search, until the common value is found or until a termination criterion is met, the iteration alternating between expanding the first record set and second record set;

in response to identifying the common value, outputting an indication identifying a candidate relationship between the first field of the first data collection and the second field of the second data collection;

using the identified candidate relationship in the creation and execution of a query to retrieve data from the data repository, the query including a join predicate specifying an equivalence or other relationship between the first field and the second field; and

transmitting the data retrieved by the query to the user device.

- 9 Claims 31 and 32 are notionally independent claims, directed towards a system or apparatus having means for performing the method of the previous claims and a computer-readable medium comprising software code adapted to perform the method of the previous claims, but the validity of these claims is wholly dependent on that of the primary claim and need not be considered independently.

The Law

- 10 Section 1(2) of the Act states:

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

- (a) A discovery, scientific theory or mathematical method;*
- (b) A literary, a dramatic, musical or artistic work or any other aesthetic creation whatsoever;*
- (c) A scheme, rule, or method for performing a mental act, playing a game or doing business, or a program for a computer;*
- (d) The presentation of information;*

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

- 11 The provisions of Section 1(2) were considered by the Court of Appeal in *Aerotel Ltd v Telco Holdings Ltd and Macrossan's Application*¹ where a four step test was set out to decide whether a claimed invention was excluded from patent protection:

¹ *Aerotel Ltd v Telco Holdings Ltd and Macrossan's Application* [2006] EWCA Civ 1371

- (1) *Properly construe the claim;*
- (2) *Identify the actual contribution;*
- (3) *Ask whether it falls solely within the excluded subject matter;*
- (4) *Check whether the actual or alleged contribution is actually technical in nature.*

12 It was stated by Jacob LJ in *Aerotel* that the test is a re-formulation of and is consistent with the previous 'technical effect approach with rider' test established in previous UK case law. Kitchen LJ noted in *HTC v Apple*² that the *Aerotel* test is followed in order to address whether the invention makes a technical contribution to the art, with the rider that novel or inventive purely excluded matter does not count as a 'technical contribution'

13 Lewison J in *AT&T/CVON*³ set out five signposts that he considered to be helpful when considering whether a computer program makes a technical contribution. Lewison LJ reformulated the signposts in *HTC v Apple* in light of the decision in *Gemstar*⁴. The signposts are:

i) Whether the claimed technical effect has a technical effect on a process which is carried on outside the computer.

ii) Whether the claimed technical effect operates at the level of the architecture of the computer; that is to say whether the effect is produced irrespective of the data being processed or the applications being run.

iii) Whether the claimed technical effect results in the computer being made to operate in a new way.

iv) Whether the program makes the computer a better computer in the sense of running more efficiently and effectively as a computer.

v) Whether the perceived problem is overcome by the claimed invention as opposed to merely being circumvented.

Analysis

14 To determine the answer to whether the claimed invention is excluded as a computer program as such, I will follow the approach set out by the Courts in *Aerotel*.

(1) Properly construe the claim

15 Both the examiner and applicant stated that the construction of the claims presents no particular challenges. This was confirmed at the hearing, and I agree that the claims do not present any notable difficulties in construing the scope of the invention.

16 I would note that while much of the discussion at the hearing and in the correspondence on file discusses the data collections being searched in the claimed

² *HTC Europe Co Ltd v Apple Inc* [2013] EWCA Civ 451

³ *AT&T Knowledge Venture/CVON Innovations v Comptroller General of Patents* [2009] EWHC 343 (Pat)

⁴ *Gemstar-TV Guide International Inc v Virgin Media Ltd* [2010] RPC 10

method as being structured in tables of rows and columns, it was highlighted at the hearing that the claimed method is not limited to use with data arranged in such tables. I think this is a fair representation, with this requirement only being introduced in claim 28. As such, the invention of claim 1 should not be viewed as being limited to use with data which is in a table format. However, I believe that it is implicit that there must be some form of structure to the data in order for the claimed method to be able to identify meaningful relationships between the various data sets.

(2) Identify the actual contribution

- 17 In their pre-hearing report of 20 April 2022, the examiner identified the contribution as ‘...the possible computational efficiency of the relation searching phase’. At the hearing, Dr Addiss argued that the present invention enabled relationships to be identified in large databases in which the use of conventional ‘brute force’ techniques to identify relationships between the datasets would be prohibitive, if not impossible, either in terms of the computational power required and/or the time it would take to complete the analysis. At the very least, it was argued that the invention enabled relationships to be identified in a practical time frame/using practical computational resources, as compared to the time/resources it would take to perform equivalent analysis with conventional approaches. It was further argued that this ultimately enabled better query results to be achieved.
- 18 While I can appreciate that the claimed invention may, as compared to conventional ‘brute force’ techniques, identify some relationships between datasets more quickly, i.e. those relating to the seed values which are used to guide the search, I am not convinced that claimed method is inherently quicker or more efficient. If one were to seek to identify all possible relationships within a large database it might even be slower than a brute force technique which simply worked through the data sets in order, as the claimed invention would likely analyse the same sets of data multiple times when different seed values were used.
- 19 It is clear from the application that the seed values are values that have either a known or suspected relationship – see, for example, lines 9-11 of page 11 of the description as it currently stands. This implies that the efficacy of the method is dependent on the accuracy or relevance of the seed values. Particularly in large data sets which cannot be searched in their entirety, the claimed method might potentially be a more efficient way of searching for relationships that exist between datasets that are connected to the seed values upon which a search is based. But I do not think that the claimed method will necessarily be better in all situations at determining any or all relationships between datasets, particularly where effective or relevant seed values are not available.
- 20 It is also within the scope of the claim to not find any relationships between the datasets, i.e. if the termination criterion is met before any relationship is noted. It may therefore be that running the claimed method provides no additional benefit to any subsequent queries. As such, I think it accurate to say that the method *may* enable better or more efficient querying of data *which is related to the seed values*.
- 21 I therefore believe that the actual contribution is:

A method of retrieving data from a data repository that comprises a number of different data collections, the method comprising performing a search process that involves receiving first and second seed values that are used to identify and retrieve first and second sets of data from first and second data collections, respectively. A common value is searched for across the first and second data sets, and if no common value is identified the search is expanded iteratively outwards using additional values from within the data sets as additional seed values, until either a common value is found or a termination criterion identified. Where a common value is identified, an indication identifying a candidate relationship between the respective fields of the first and second data collections is outputted which can be used in the creation and execution of subsequent queries to retrieve data related to the seed values in a potentially more efficient manner.

(3) and (4) Ask whether the contribution falls solely within the excluded subject matter, and check whether the actual or alleged contribution is actually technical in nature

- 22 At the hearing, it was argued that the *AT&T* signposts (ii), (iii) and (iv) were most relevant to the invention.
- 23 In respect of signpost (ii), whether the claimed invention works at the level of the architecture of the computer, it was argued that technological developments are such that the traditional view of a ‘computer’ as being a stand-alone, monolithic device was no longer appropriate. Rather, computer systems are now often much more distributed and network-based and include computer devices interacting with remotely stored data and remotely provided computing resources via interconnected infrastructure components. These infrastructure components provide an architectural foundation upon which modern computer systems are constructed. As such, the concept of computer architecture should encompass the implementation of and interaction between the infrastructure components that support end-user applications. Improvements made to distributed computer architectures and the infrastructure components that allow them to operate should be considered to be operating at the level of the architecture of a computer.
- 24 The present invention is related to data storage and data access architectures, which, according to Dr Addiss, underpin how modern distributed computer systems are built and developed. The data repository of the present invention, which utilises the claimed method to identify relationships between different data sets, allows efficient centralised access to data from various data sources via the specified search and query process. It was argued that this process does not operate at the application level and is designed to operate independently of the nature of the processed data and the applications being run which are accessing the data.
- 25 I do not find this argument persuasive. I believe it is correct to consider that ‘a computer’ may potentially comprise a system or network of computers – this point was addressed by Birss J. in *Lantana v Comptroller-General Patents*⁵ in relation to signpost (i), and more generally I consider it reasonable to accept that the concept of ‘a computer’ is not limited in this day and age to a single, monolithic device.

⁵ *Lantana v Comptroller-General Patents* [2013] EWHC 2673 (Pat)

However, I think this argument conflates what is meant by 'architecture' in signpost (ii) and the broader meaning of 'architecture' as it relates to the structures and components which might make up a computer network. Central data repositories and the like are obviously part of the architecture of a computer network in general terms, but signpost (ii) is specifically concerned with those components and operations which operate at the most fundamental levels of a computer, and which operate in the same way irrespective of the particular functions or applications the computer is tasked to perform.

- 26 It was argued that the claimed method can work with any data, but I am not convinced that that is the case as that data must have some form of appropriate structure. Furthermore, all that the claimed method enables a computer to do is to potentially better identify relationships between data sets. Fundamentally, it is about how the data itself is structured to facilitate querying the data. For something to be considered to be operating at the architectural level of a computer, it must be capable of assisting a computer to perform any higher level or application-based task. I do not believe that the claimed invention has the utility required to enable a computer to perform a broad range of different and diverse tasks better as would be required for it to be considered to operate at the architectural level of the computer. It may enable a specific task to be performed better by the computer, and indeed at a practical level there may be some further benefits that can be obtained if the method enables queries to be more efficiently processed, but it does not assist the computer at the level of generality that I think is required for signpost (ii) to be met.
- 27 In respect of signpost (iii), whether the claimed technical effect results in the computer operating in a new way, it was again argued that the computer should be considered to be a distributed system formed from a number of different infrastructure parts that enable the system to operate. The provision of additional infrastructure into the distributed system to enable better tailored queries to be made to distinct data sources within the system was argued to result in the computer system itself operating in a different manner, allowing the centralised querying of data from a plurality of data sources that wouldn't otherwise be possible. Effectively the claimed invention sits between various aspects of the computer infrastructure and facilitates a new form of operation between those aspects.
- 28 I am not convinced by this argument. I can see nothing that implies that the computer, distributed or otherwise, is operating in a new way – the centralised querying of data is well known in general terms. Rather, it would appear that there is a computer program that is potentially operating in a new or better way, but upon a computer that is operating in a normal and conventional fashion. The computer program merely relates to the way the relationships between the data are defined so that future queries can take advantage of those relationships. The computer itself operates conventionally.
- 29 In relation to signpost (iv), whether the computer program makes the computer a better computer in the sense of it running more efficiently or effectively, it was again argued that 'a computer' should be interpreted to include a distributed system as per the present invention. It was further argued that the claimed method enables a particularly efficient identification of data relationships between data from different data sources which in turn allows for efficient querying of that data by client devices, even for very large data sets. The computer system as a whole, in the sense of the

distributed data retrieval system, works better as a result of the present invention as the data can be queried and accessed more efficiently.

- 30 I am not convinced that this argument meets the requirements of signpost (iv) which is concerned with whether the computer as a whole can be considered to run more efficiently and effectively as a result of a computer program being run. This does not appear to be the case with the present invention, in that only the querying function may potentially be considered more efficient or effective, on those occasions where the method has identified any suitable relationships. The invention may be a better computer program, which may therefore require fewer computational resources or reach its conclusion in a shorter time, but it fails to provide the required improvements to the computer itself (be it distributed or otherwise) at the general level required to satisfy signpost (iv).
- 31 Signposts (i) and (v) were not considered in detail at the hearing but were briefly discussed for completeness. It was not argued that the invention has a technical effect on a process carried out outside of the computer, as per signpost (i), and I can see no evidence of any such effect. The program identifies relationships between values stored in database systems and as such relates to the structuring of data. This takes place entirely within the computer and there is no effect on a process outside of the computer.
- 32 For signpost (v) it was argued that the invention solves a technical problem and does not merely circumvent the problem. However, I am not convinced that this is the case. The primary argument put forward at the hearing was that some modern data repositories are simply too large for conventional 'brute force' methods to map out, or form a schema for, in a reasonable time frame and/or using reasonably accessible computational resources. It was argued that the claimed method enabled such data repositories to be efficiently searched in a way that conventional methods do not. I do not think that the claimed method solves this problem, rather it effectively prioritises the available time and resources on those areas of the data repository where it is suspected there is the best chance of finding relationships. It does not, however, enable a full analysis of a large data repository to be completed more quickly or efficiently. Indeed, as previously discussed it would appear that the claimed method may be less efficient at covering the entire data repository than a 'brute force' method that simply worked through all of the data methodically. As such, the claimed invention appears to be a form of prioritisation that acts to minimise the impact of the problem as it was set out in the hearing. It does not solve the perceived problem in any technical sense, for example by improving the technical infrastructure of the data repositories or the way the data is stored in the repositories, but rather circumvents it by changing the way the data is organised. I do not think signpost (v) therefore provides any clear indication that the claimed invention might be patentable.
- 33 While it is worth noting that the signposts are not a definitive test and are only useful aids against which to consider the patentability of an invention, I do not believe that any of the signposts provide any indication that the claimed invention is more than a computer program as such.
- 34 In addition to discussions around the *AT&T* signposts, a number of decisions by the European Patent Office's Technical Board of Appeal were also brought to my

attention in the skeleton argument provided ahead of the hearing. It was made clear at the hearing that the primary reason for highlighting these decisions was to draw attention to where the EPO Board of Appeal had suggested that improved search methods could be considered technical. As I noted at the hearing, EPO decisions are persuasive but not binding and, particularly with such cases, care must be taken when drawing analogies. Comments must be considered in the context of the particular facts of the case, and it is not necessarily the case that such comments can readily be separated from that context and still be meaningfully applied. Rather, such comments need to be considered in light of the wider decision and the invention with which it is concerned.

- 35 Paragraph 9 of T1351/04, which was referred to during prosecution, was highlighted in the hearing as stating that an index file was deemed to be a technical means on the basis that the management information which it contains determines the way in which a computer searches information, which is considered to be a technical task. As such, claim 1 of the application in T1351/04, which related to a method of generating an index file, was deemed to have a technical character. It was argued that the method of the present invention identifies relationships, in this case between datasets in a database, with those identified relationships then being used to guide and focus queries in a similar way to the operation of an index file. As such, claim 1 of the present application should be considered to show a relevant technical effect.
- 36 I do not believe that the current claimed invention can be deemed to be on all fours with that of the application with which T1351/04 is concerned. The latter effectively sets out a methodology by which a computer can more efficiently search files. By comparison, the current invention is merely concerned with identifying possible relationships within datasets, but the actual searching which is performed thereafter is wholly conventional.
- 37 Three further decisions, T1849/17, T0643/00 and T1569/05, were also referred to, primarily with a view to reinforcing the view that search methods are a technical process. Specific attention was drawn in the hearing to paragraph 7.2 of T1849/17, which was argued to support the view that considerations aimed at improving the searching and retrieval of data in a more efficient or faster manner can be considered to be technical in nature.
- 38 Beyond supporting the general proposition that search methods may be patentable, which is not a point with which I disagree, I do not believe that these decisions provide any great assistance. The question that must be answered is not whether methods of searching may be patentable but rather if the claimed invention is a computer program as such, and it is important not to conflate these two questions. This question must be decided on the facts of the particular case, and the inventions with which the EPO decisions are concerned do not appear to have any great relevance to the invention with which this decision is concerned. As such, I do not believe that the EPO decisions provide any particularly helpful guidance, and in particular do not provide any useful evidence that the contribution of the claimed invention is more than a computer program as such.
- 39 To summarise, there is no suggestion that the contribution provides a technical effect that acts on a process outside of the computer. I do not believe that the technical effect can be considered to operate at the level of the architecture of a computer,

irrespective of whether you consider 'a computer' broadly as an interconnected network of devices. Nor do I think that the computer is made to operate in a new way or run more efficiently or effectively as a result of the running of the computer program. Moreover the perceived problem is not being solved in any technical sense. Beyond the *AT&T* signposts and bearing in mind the decisions which were highlighted at the hearing, whilst the claimed invention may be a better computer program, I can see nothing which indicates to me it can be considered more than a computer program as such. Fundamentally the invention provides an algorithm for identifying relationships between data fields in records which a query can then make use of. It is about how the data is structured at a high level. The querying itself operates in a conventional manner once these relationships have been identified. There is no technical contribution but rather the contribution lies wholly in the excluded field of a program for a computer as such.

Conclusion

- 40 In conclusion, I have found that the claimed invention lies solely in the excluded field of a program for a computer as such and therefore does not comply with the requirements of sections 1(1)(d) and 1(2)(c) of the Act. I therefore refuse the application under section 18(3).

Appeal

- 41 Any appeal must be lodged within 28 days after the date of this decision.

B Micklewright

Deputy Director, acting for the Comptroller