Global Yellow Pages Limited *v* Promedia Directories Pte Ltd and another suit [2013] SGHC 111

Case Number	: Suits Nos 913 and 914 of 2009 (Registrar's Appeals Nos 421 and 422 of 2012)
Decision Date	: 22 May 2013
Tribunal/Court	: High Court
Coram	: Lee Seiu Kin J
Counsel Name(s)	: Karen Teo, Adeline Chung and Han Hsien Fei (TSMP Law Corporation) for the plaintiff in Suit No 913 of 2009 and Suit No 914 of 2009; G Radakrishnan (Infinitus Law Corporation) for the defendant in Suit No 913 of 2009; Zhulkarnain Abdul Rahim and Diyanah Binte Baharudin (Rodyk & Davidson LLP) for the defendant in Suit No 914 of 2009.
Parties	: Global Yellow Pages Limited — Promedia Directories Pte Ltd

Civil Procedure – Discovery of documents – E-discovery

22 May 2013

Lee Seiu Kin J:

Introduction

1 While modern technology has solved numerous problems and greatly facilitated daily activities, it has also given rise to new issues. Today, information is transmitted around the globe literally at the speed of light and electronic content can easily be created, altered and disseminated. The sheer volume of electronic information, as well as the difficulty of accessing some types of electronic information, presents considerable practical challenges in the area of discovery in litigation. The law requires all relevant documents to be disclosed except where such disclosure is not necessary either for disposing fairly of the matter or for saving costs, or where the documents are privileged. However, in many circumstances today it would be enormously impractical for a party to visually inspect each and every electronic document to determine whether it should be disclosed. It must be borne in mind that the common law procedures for discovery were developed long before the era of the computer and Internet. As technology develops, so must the law adapt in appropriate fashion.

2 These two appeals, *viz*, Registrar's Appeal No 421 of 2012 ("RA 421") in Suit No 913 of 2009 ("Suit 913") and RA No 422 of 2012 ("RA 422") in Suit No 914 of 2009 ("Suit 914"), were appeals against the decision of the assistant registrar ("AR") ordering that searches using particular keywords be conducted on various electronic devices.

3 After considering the parties' submissions and oral arguments, I dismissed both appeals except that I made two minor variations to the AR's orders (see [23] below). I now set out my reasons for doing so.

The facts

The parties

4 The plaintiff in both suits, Global Yellow Pages Limited ("GYP"), is a company incorporated in

Singapore and listed on the Singapore Exchange. It carries on the business of publishing directories and providing classified directory advertising and associated products and services.

5 The defendant in Suit 913, Promedia Directories Pte Ltd ("Promedia"), is a company incorporated in Singapore and carries on the business of publishing directories.

6 The defendant in Suit 914, Streetdirectory Pte Ltd ("Streetdirectory"), is also a company incorporated in Singapore. GYP averred that Streetdirectory carries on the business of publishing directories. Streetdirectory averred that it is in the business of developing location-based software and solutions, including specialised skills in the areas of "geomatics", wireless communication, global positioning system tracking, vehicle navigation and mobile applications.

The pleadings

GYP's claim

7 GYP's pleadings in Suit 913 and Suit 914 were largely similar. What I set out in this section should therefore be taken as a reflection of GYP's pleadings as against both Promedia and Streetdirectory unless otherwise indicated.

8 GYP publishes a series of directories which contain lists of companies and businesses and their respective contact details (such as names, telephone numbers, addresses, facsimile numbers, branches or subsidiaries, and company registration numbers or licence numbers). These directories consist of (a) printed directories which are published in annual editions ("Printed Directories"), and (b) an electronic directory ("the Online Directory").

9 The Online Directory is updated daily and is known as the "Internet Yellow Pages". A person who wishes to access information on the Online Directory may enter search terms at http://www.yellowpages.com.sg ("the Search Engine"). The Search Engine produces results which are similar to what is found in the Printed Directories, unless the Online Directory has since been updated.

10 GYP averred that it (and/or its employees) was the author of the following original works, and that GYP was at all material times the owner of the copyright which subsisted in the following works (collectively, "GYP's Works"): [note: 1]

(a) The 2003/04, 2004/05, 2005/06, 2006/07, 2007/08, 2008/09 and 2009/10 editions of the Printed Directories, which were compilations by reason of the selection and arrangement of their contents, thereby constituting intellectual creations.

(b) The Online Directory, which was a compilation by reason of the selection and arrangement of its contents, thereby constituting an intellectual creation.

(c) Subscriber information, which had been verified, enhanced, arranged and classified by GYP from information provided by Singapore Telecommunications Limited and Starhub Limited, as found in the abovesaid editions of the Printed Directories and the Online Directory.

In respect of its claim against Streetdirectory, GYP also averred that it was the owner of the copyright which subsisted in the 1998/99, 1999/2000, 2000/01, 2001/02, and 2002/03 editions of the Printed Directories. These editions of the Printed Directories were not pleaded in GYP's claim against Promedia.

11 GYP averred that from or about 2003, Promedia had infringed the copyright in GYP's Works or parts thereof by reproducing or authorising its reproduction, without GYP's licence or consent, in Promedia's own directories (collectively, "Promedia's Directories"), namely (a) Promedia's online directory at http://www.thegreenbook.com ("Promedia's Website"), (b) the 2003 to 2009 editions of Promedia's printed directories (known as "The Green Book"), and (c) the 2003 to 2009 editions of The Green Book CD-Rom. [note: 2] GYP made similar allegations against Streetdirectory, except that the alleged reproduction was contained in Streetdirectory's online directory at http://www.streetdirectory.com ("Streetdirectory's Website"). [note: 3]

12 GYP alleged that this reproduction could be seen from: [note: 4]

(a) Substantial similarities between listings in, on the one hand, GYP's Printed Directories and Online Directory, and, on the other hand, Promedia's Directories (or Streetdirectory's Website).

(b) The presence of fictitious subscriber listings ("Seeds") "planted" by GYP throughout its Printed Directories and Online Directory in Promedia's Directories (or Streetdirectory's Website).

The Seeds were fictitious company or individual names with addresses and telephone numbers which did not belong to these fictitious companies or individuals. [note: 5]

Promedia's defence and counterclaim

13 Promedia averred that GYP's Works were not original literary works in which copyright could or did subsist. [note: 6]

14 Promedia further averred that even if there was copyright in GYP's Works, it had not infringed such copyright. [note: 7] Promedia alleged that over the past 31 years, it had independently created its own database of subscriber information or data. [note: 8] It went on to note that because telephone directories are fact-based works, there will inevitably be similarity in directories published by different publishers because the facts will be similar, and that such similarity could not amount to copyright infringement. [note: 9]

15 Promedia averred that the presence of the Seeds in Promedia's Website and in The Green Book CD-Rom was negligible and minimal and therefore did not amount to substantial copying such as to constitute copyright infringement. [note: 10]

16 Promedia averred that even if it had infringed GYP's copyright, it could rely on the defence of fair dealing under s 35 of the Copyright Act (Cap 63, 2006 Rev Ed). [note: 11]

17 Promedia's defence also contained various other averments, some of which were as follows: (a) Promedia could rely on a defence of public interest; (b) GYP was guilty of laches, acquiescence and/or delay; and (c) Promedia was an "innocent infringer" because, at the time of the infringement, it was not aware and had no reasonable grounds for suspecting that GYP had copyright in the relevant compilations, data or listings. [note: 12]

18 Promedia also counterclaimed for loss and damage as a result of GYP's groundless threats of copyright infringement. [note: 13] GYP's defence was a denial of Promedia's allegations in the counterclaim. [note: 14]

Streetdirectory's defence

19 A large part of Streetdirectory's defence consisted of the non-admission of GYP's pleadings.

20 In addition, Streetdirectory denied that it had infringed GYP's alleged copyright by reproducing the whole of GYP's Works or a substantial part thereof in a material form in Streetdirectory's Website. [note: 15] Streetdirectory averred that each entry in Streetdirectory's Website was a collection of factual information related to a business entity, including the name, address, telephone number, fax number, email address and website of that business entity, and that GYP and itself both relied on the same factual information. [note: 16]

The proceedings before the Assistant Registrar

21 On 17 February 2012, GYP applied for discovery against Promedia and Streetdirectory by way of, respectively, summons no 773 of 2012 ("SUM 773") in Suit 914 and summons no 774 of 2012 ("SUM 774") in Suit 913. Both applications prayed for, *inter alia*, the following orders:

(a) That there be discovery of various electronic documents in the possession, custody or power of Promedia or Streetdirectory pursuant to paras 43C(4) and 43C(6) of the Supreme Court Practice Directions in force at that date; and

(b) That such discovery be carried out in accordance with the electronic discovery protocols which were annexed to the applications ("the Protocols").

22 What was material for the purpose of the appeals before me was para 1(*d*) of the Protocols, which contained several keywords which would be used in the conduct of a search of specified devices. The AR granted an order that several keywords be used in the conduct of the said search. Promedia and Streetdirectory both appealed against the AR's decision by way of, respectively, RA 421 and RA 422. For the purpose of these appeals, the relevant keywords which were ordered by the AR to be used were as follows:

Keyword	**Promedia**	**Streetdirectory**
"seed", "seeds"	Yes	Yes
"copy" "follow"	Yes	Yes
"delete", "destroy", "erase", "wipe out" "remove"	Yes	Yes
"directories"	Yes	Yes
"classifications"	Yes	Yes
"business listings"	Yes	Yes
"Global Yellow Pages", "GP", "GYP", "Yellow Pages", "YP", "YPS"	Yes	Yes

"[name of each custodian/employee materially involved in creating, authoring and/or compiling the directories]"	No	Yes

The appeals were in respect of the same keywords, except that Streetdirectory also appealed against the AR's order in respect of one additional keyword: "[name of each custodian/employee materially involved in creating, authoring and/or compiling the directories]".

23 I dismissed both appeals subject to the following variations:

(a) The search for the term "seed" or "seeds" on the relevant custodians was to be limited to the period from 17 September 2010 to 6 November 2012, with liberty to apply if GYP was able to prove that this expression entered into the common vocabulary at an earlier date.

(b) The purported keyword "[name of each custodian/employee materially involved in creating, authoring and/or compiling the directories]" (see [22] above) was to be deleted from the first column of the table pertaining to the keyword search contained in the Protocols. Properly construed, this purported keyword was not a search term but was instead a limitation on the scope of the search to be conducted. This limitation was already contained in the second column of the table pertaining to the keyword search.

I also ordered that costs be in the cause and fixed the quantum of costs and disbursements at $1,000 in each suit.

Discovery and its origins

24 Pursuant to O 24 of the Rules of Court (Cap 322, R 5, 2006 Rev Ed), the court may order a party to a suit to give discovery of all documents which are, or have been, in his possession, custody or power that are relevant to the issues in dispute if such discovery is necessary either for disposing fairly of the cause or matter or for saving costs.

25 In *Fermin Aldabe v Standard Chartered Bank* [2009] SGHC 194, the process of discovery was outlined by Yeong Zee Kin SAR as follows:

> 27 Without delving into the case law relating to discovery, I propose first to set out briefly the classical sequence in which discovery and inspection takes place under Order 24. Discovery of documents is given by enumerating them in a list of documents, the completeness of which is to be verified by affidavit. The list typically contains a notice stating the time and place at which the party served with the notice may inspect the documents referred to in the list. The list and affidavit verifying are served on the other parties in the action. At the appointed time and place, the enumerated documents are produced for physical inspection and the inspecting party is entitled to take a copy of the inspected documents during inspection.
>
> 28 Hence, inspection and the taking of copies occurs concurrently in the classical sequence. However, there is a common practice for parties to agree that copies of all documents enumerated in each party's list of documents be exchanged first and for physical inspection of documents to be deferred. After copies of documents have been exchanged, a party is still entitled to request for inspection of specified documents pursuant to the agreement to defer physical inspection. ...

26 What we now know as discovery can be traced back a long way in history. AE Randall (ed), *Commentaries on Equity Jurisprudence by Justice Story* (Sweet & Maxwell, 1920, 3rd Ed) suggested (at p 628, §1487) that "the probable origin of [proceedings for discovery] in our courts of equity" was to be found in Roman law. Edward Bray, *The Principles and Practice of Discovery 1885* (Legal Books, 1985 Reprint) noted (at p 5) that:

> The jurisdiction to give discovery in chancery to sustain an action at law seems to date back as far as Henry VI [in the 15th century]. ...

27 By the late eighteenth century, the concept of discovery had become firmly entrenched in the courts of equity. *The Supreme Court Practice 1976 Volume 1* (IH Jacob gen ed) (Sweet & Maxwell, 1975) stated (at para 24/1/1):

> **The History of Discovery.**—By the end of the eighteenth century, the Courts of Equity (the Court of Chancery and the Court of Exchequer in its equitable jurisdiction) had evolved a method of proof to which the general name "discovery" was given, and which comprised: (*i*) *discovery of deeds and documents*, by which a person could be compelled to produce for inspection deeds or documents relevant to a dispute which were in his possession or power; this procedure was the foundation of discovery in the modern sense as dealt with by this Order; (ii) *discovery of facts* by which a person might be ordered to answer as to the existence of some fact within his knowledge and relevant to a dispute; this form of discovery was the origin of interrogatories ...
>
> The Common Law Courts did not have discovery as such, although there were restricted, and very technical, methods of obtaining inspections of documents ... But in cases where these methods were not available, a party to a Common Law action who wished to have discovery filed a bill in equity for this relief only, the action being adjourned meanwhile.
>
> [emphasis in original]

28 Paul Matthews & Hodge M Malek QC, *Disclosure* (Sweet & Maxwell, 2012, 4th Ed) ("*Disclosure*") observed (at para 1.11):

> The origins of discovery in English law are obscure, but they appear to lie in the procedures of the civilian courts, such as the ecclesiastical courts. But from an early date similar techniques were also developed and employed in the various courts of equity, of which the most important was the Court of Chancery. By the late eighteenth century, the plaintiff's bill of complaint in a Chancery suit invariably had three parts: allegations of fact (stating part), evidence (charging part) and interrogatories to the defendant (interrogating part). Thus the bill included a discovery aspect from the outset. The plaintiff could also obtain discovery and production of documents from the defendant by a separate "bill of discovery". ...

Discovery and electronic documents

The problem: quantity and retrieval of electronic information

29 Discovery, having its relatively modern origins in the courts of equity, is meant to serve the ends of justice and a fair trial by increasing the likelihood of the court resolving the dispute on the basis of facts which represent or which are reasonably proximate to the truth: see *Davies v Eli Lilly & Co and Others* [1987] 1 WLR 428 at 431H–432B. If taken beyond its proper bounds, however, the process of discovery may conduce towards injustice. The editors of *Disclosure* observed (at para 1.03):

> Disclosure is not without its disadvantages. The principal one is that disclosure can be an expensive and burdensome process. The courts are generally alert to the danger of oppressive disclosure and inappropriate requests for wide-ranging disclosure are not infrequently dismissed for being not necessary for the fair disposal of litigation. The burden can not only fall on the party giving disclosure, but also on an opposing party presented with a mass of documentation of marginal relevance. *In such a case disclosure can, far from clarifying the issues, operate as a cloud.* ... [emphasis added]

30 The risk of oppression may be heightened if the documents sought consist of electronic documents. In the foreword to *Disclosure*, Etherton LJ observed extra-judicially that:

> ... [L]ike so many other forensic tools for a fair trial disclosure requires careful control because it can itself become an instrument of oppression both for those who must give disclosure and those to whom the disclosure is made. That oppression may take the form of the quantity of disclosure demanded or given, the time and effort involved for all parties, and cost. Those matters, and particularly cost, may themselves be a practical impediment to access to justice. ... *[This] is a point of particular significance in relation to the disclosure of electronic documents*. [emphasis added]

31 In *Breezeway Overseas Ltd and another v UBS AG and others* [2012] 4 SLR 1035 ("*Breezeway*"), at [20], I had observed:

> The perennial tension in the law of civil procedure, *viz*, the attempt to achieve both justice and efficiency, comes to the forefront in the discovery process. On the one hand, it is *ex hypothesi* in the interest of justice that all relevant material is discovered, while on the other, there is a pressing need to ensure efficiency lest injustice be occasioned through the well-meaning but disproportionate attempt to ensure that all relevant material is disclosed.

Jacob LJ had famously cautioned against a "leave no stone unturned" approach in the search for "perfect justice" in every case as this would actually defeat justice: *Nichia Corporation v Argos Ltd* [2007] Bus LR 1753; [2007] EWCA Civ 741 ("*Nichia*"), at [50]–[51].

32 The reason is that, as some commentators put it, "an evolutionary burst in writing technology" has led to an "explosive growth of information" and "information inflation" (see George L Paul & Jason R Baron, "Information Inflation: Can the Legal System Adapt?" (2007) 13 Richmond Journal of Law & Technology 10 ("*Paul & Baron*") at paras 9, 11 and 14) [note: 17]. In addition to writing technology contributing to increased authorship, storage technology has exacerbated the problem exponentially by making it easy to generate multiple copies and feeding the human propensity to hoard soft copies indefinitely. I described this problem in *Sanae Achar v Sci-Gen Ltd* [2011] 3 SLR 967 ("*Sanae Achar*") in the following manner:

> 12 The introduction of the [Supreme Court Practice Direction No 3 of 2009 ("the e-Discovery PD") in 2009] was a response to *the increasing tendency for documents to be generated and held electronically.* In my view, its introduction was timely, given *the unprecedented volume of documents which are created and stored electronically today* (attributed, in part, to the ease at which multiple copies of the same document, especially e-mails, are stored at multiple locations, *eg*, personal computers, servers, network drives, and other assorted backup media, for indefinite periods of time, and at relatively low costs), the relative ease of duplicating such documents (by way of illustration, the same e-mail may be sent to multiple recipients, who may reply to one or more recipients on the e-mail thread, copying in other recipients, or forwarding the message on to others), the often haphazard manner in which electronic documents are stored, the different

document retention policies of parties (some may routinely delete electronic documents to maximise the use of storage capacity whereas others may retain records of all electronic documents), the existence of metadata information, and the fact that it is more difficult to completely dispose of electronically stored documents than printed ones (it is common for residual traces of an electronic document to remain on a computer's storage system, despite deletion of the same from the user's active data).

1 3 *With technology fuelling an unprecedented explosion of the volume of discoverable documents and the ease of their duplication, it is not surprising that the traditional manner in which discovery has been carried out is proving increasingly inefficient in achieving the purposes for which the discovery process was developed.* Discovery was originally an equitable remedy premised on the idea that it was unconscionable for a party to conceal evidence material to a fair conclusion. Its aim was that of enabling parties to acquire information which is material to their case but in the possession, power or custody of their opponents. Discovery was directed towards the just and fair adjudication of an action as it ensured that all material and relevant facts were placed before the court for its consideration. ...

[emphasis added]

33 Consequently, we often do not know what we have and are therefore hard pressed to give. This concern has been articulated by Senior Master Whitaker in *Gavin Goodale & Ors v The Ministry of Justice & Ors* [2010] EWHC B40 (QB) ("*Goodale*") in the following manner:

4 ... in the case of paper disclosure, parties usually know what paper they have. Often the problem is merely locating it physically and going through it to produce the documents required by the standard disclosure test. The problem with ESI is that, … parties often do not know how much ESI they have, or where it is. They might have a idea as to which servers it is on or which personal computers it is on, or which back-up tapes it is on, but without a great deal more information, it is very difficult for them to know how much documentation will be revealed by searches of the media on which their ESI is stored and how much it is going to cost to search it and what the end result is going to be. A further issue might be that not all forms of ESI are searchable. …

34 The interests of efficiency requires that a case gets to trial as soon as possible with the best set of documents that can be amassed to assist in arriving at a decision on the merits. Commercial entities look for finality as it brings an end to disputes and, win or lose, they can put the dispute behind them, write off bad debts and get on with their business. Efficiency seeks to cull the volume of documents to be disclosed and it employs the scythe of proportionality and economy. The ultimate goal is to ensure that burgeoning volumes of discoverable documents do not translate into burgeoning legal costs that may prevent all but those litigants with the deepest pockets from seeing their cases tried on the merits. The Holy Grail is to arrive at a set of documents of the right size containing all relevant documents without expenditure of disproportionate costs.

35 The ills of burgeoning legal costs are illustrated by the following passage from *Nichia*, where Jacob LJ observed as follows:

46 ... It is wrong just to disclose a mass of background documents which do not really take the case one way or another. And there is a real vice in doing so: it compels the mass reading by the lawyers on the other side, and is followed usually by the importation of the documents into the whole case thereafter—hence trial bundles most of which are never looked at.

47 Now it might be suggested that it is cheaper to make this sort of mass disclosure than to consider the documents with some care to decide whether they should be disclosed. And at that stage it might be cheaper—just run it all through the photocopier or CD maker—especially since doing so is an allowable cost. But that is not the point. *For it is the downstream costs caused by over-disclosure which so often are so substantial and so pointless.* It can even be said, in cases of massive over-disclosure, that there is a real risk that the really important documents will get overlooked—where does a wise man hide a leaf?

[emphasis added]

Technology provides a solution to the problem

36 While technology has created the problem, it also presents us with a means to alleviate or solve the problem, without which the litigation process could come to a grinding halt for all but those with the deepest pockets. In *Sanae Achar* (cited at [32] above), I had advocated the use of technology as a means of coping with the burgeoning volume of discoverable documents:

14 One way to cope with the burgeoning volume of discoverable documents is to rely on technology itself. … Running simple keyword searches using easy-to-use desktop search engines would suffice. It is also easier to manage and organise electronically stored documents, especially where printed copies of such documents run into tomes and cartons. The e-Discovery PD recognises the tremendous potential of technology in modernising the discovery process. Thus, it encourages the exchange and supply of copies of discoverable electronic documents in soft copy by creating a framework for the inspection and discovery of electronically stored documents within boundaries established by existing legal principles. ...

37 Similarly, Joel Greer, "Practical Considerations Regarding Certain Aspects of Electronic Discovery" (2009) 12(2) International Arbitration Law Review 11 ("*Greer*") observed as follows (at pp 14–15):

In both US judicial opinions and the secondary literature, there has been considerable discussion regarding use of various automated computer search tools to enable litigants and their counsel to examine large volumes of ESI in a cost-effective manner. It is now accepted that use of such tools is an extremely valuable, and even necessary, part of the ESI review process given that the volume of ESI in parties' possession may make it prohibitively expensive, or physically impossible, for reviewers to conduct document searches manually. ...

38 In Singapore, the most common method of retrieving and analysing electronic documents that has come before the courts is the familiar keyword search utilising Boolean operators. Keyword searches are widely available today and are familiar to most lawyers. The reason for this is posited by *Paul & Baron* as follows:

[37] ... the *status quo* for the legal profession is to use "keywords," without more, to ferret out electronically-stored information in large corporate and institutional databases. The legal profession has adopted keyword searching in light of its longtime familiarity with its use in connection with the offerings of the major online legal retrieval services, ...

39 Before I go on to set out my views on keyword searches, which constituted the subject-matter of the two appeals before me, I will briefly set out the courts' approach in the use of search technologies. Generally, there are two categories of search technologies: the familiar keyword searches and concept searches, a class of search technology that is gaining increasing prominence.

See *Paul & Baron* at paras 42–43 for a handy overview of differences between keyword and concept search technologies.

40 Alternatives to search technology like predictive coding – designed to be used iteratively in tandem with human review – may in future find increasing prominence. Search technology cannot be the only tool that lawyers utilise to tame the burgeoning beast. Apart from search technology, modern document review platforms have a fairly standard set of document review and management tools to remove duplicates, present emails in threaded conversations, cluster conceptually similar documents together for review, *etc*, see Daniel B Garrie & David Harvey, *E-discovery in New Zealand: the impact of the new High Court Rules* (2012) 31(3) CJQ 305 ("*Garrie & Harvey*") at pp 315-316 for a more detailed description of these technical terms.

41 In the interests of promoting efficiency in civil procedure, our courts do embrace and encourage the adoption of modern search technologies and document review and management tools. Our courts adopt a technology-neutral approach, showing no preference for any particular type of search technology or document review and management tool. However, the march of technology means that there are and will be better tools available for use by lawyers or litigation support professionals in the future. Indubitably, the law and practice will evolve in tandem with such developments to refine and improve our discovery process so that it remains a force for justice and not injustice. The adversarial system will play its role in ensuring that advocates adopt trustworthy technology. As technology evolves to perform more efficiently and accurately, litigants will benefit from them. The courts have encouraged lawyers to harness search technologies and document review and management tools to bring efficiency to civil litigation practice, especially during discovery: these tools increase the efficiency and effectiveness of lawyers and, if used adroitly, will lower the costs of litigation, not only during discovery but downstream as well, *eg*, during the preparation of affidavits of evidence-in-chief and witness preparation for trial.

Keyword searches

42 Today, the most common method of e-discovery in Singapore appears to be keyword searches. As most who have used keyword searches have experienced, not all search results result in retrieving relevant documents. This limitation of keyword searches is succinctly explained in *Garrie & Harvey* at p 314 as follows:

> Key words operate by returning results that contain a word or combination of words. However, parties should not infer that the delivery of a result means that the documents are relevant because the key word used can have different meanings based on how the term is used. Thus, when parties elect to locate documents by way of keywords, the selection of the appropriate terms is critical.

43 But that is not to say that keyword searches should not be used or that ocular review is always better. In *Breezeway* (cited at [31] above), I had made the observation that both keyword searches and ocular review are not without shortcomings:

> 24 Keyword searches are potentially both over- and under-inclusive. As any user of search engines would realise, false positives and false negatives are an inevitable result of attempting to identify relevant material through keyword searches. This should be contrasted with ocular review, which could theoretically ensure zero gaps in the identification of relevant material. However, when a large number of documents have to be reviewed, discrepancies inevitably arise due to fatigue and variances in each reviewer's subjective appreciation of the issues in dispute and threshold for relevance. When the documents to be reviewed exceed a certain volume,

accurate ocular review becomes prohibitively costly and impracticable.

25 Imperfect as they are, therefore, keyword searches present a practical trade-off between achieving a theoretically complete set of relevant material and keeping costs proportionate to the value of the claim. As part of the e-discovery process of identifying relevant material, keyword searches provide a pragmatic solution where the costs of ocular review would be way out of proportion to the stakes in the case.

An iterative process

44 In my view, the key to unlocking the full potential of keyword searches (bearing in mind the difficulties outlined above) is by way of adopting an *iterative* process. This will help to mitigate the difficulties and/or inherent limitations in keyword searches. *Greer* states as follows (at p 16):

> ... many commentators recommend an iterative process in which litigants and their counsel develop keyword search terms and Boolean operators together through negotiation. A simple example of this process would be as follows: a party receiving requests for ESI document disclosure drafts and proposes advanced keyword searches to its adversary, which reviews the proposed keyword searches and offers amendments, whereupon the parties endeavor to reach consensus on the search terms to be used. The transparency between parties and counsel regarding search terms does not (nor is it intended to) prevent disputes from arising. Nevertheless it is a necessary basis for the parties themselves to be able to negotiate search protocols and try to resolve disagreements without reference to a tribunal.
>
> Is such iterative collaboration superior to a process where, for example, each party unilaterally chooses keywords (and other search criteria) and simply discloses its choices to the other? For several reasons, it is suggested that the answer is yes. First, as noted above, keyword searching is inherently imperfect, and each party and its counsel can directly affect the margin of error through their choice (and exclusion) of keywords and other search criteria. As a result, parties and counsel have an interest in critically examining their adversary's selection of keywords and other search criteria. The iterative collaboration process outlined above would best ensure that parties and counsel do so consistently.
>
> Secondly, as use of keyword searching becomes more widespread in arbitration, even under the unilateral selection-and-disclosure approach it is to be expected that parties and counsel will anyway scrutinise opponents' keywords with greater frequency. Because more disagreements are likely to arise as a result, having a process ex ante for parties to manage differences co-operatively should help reduce cost and delay. Further, collaborative involvement by both sides in developing keyword searches gives each party a stake in the outcome, making it more difficult for either side to challenge agreed search protocols later and thus lessening the chance that disputes will occur subsequently that require reference to a tribunal. Lastly, because it includes active input from diverse and opposing perspectives, the give-and-take among parties and counsel regarding keywords and other search criteria can be expected to generate better search capability (i.e. retrieve more relevant documents), which, again, is in both parties' interest.

4 5 *Paul & Baron* have optimistically observed that the iterative process may result in "virtuous cycles":

> [51] ... in response to the problem of searching large data sets, one can expect "virtuous cycles" in the form of iterative feedback loops where multiple, iterative meet and confer sessions occur for information exchange and discussion of issues to research, negotiate, and agree.

46 This iterative process has been endorsed by the courts. To take one local example, Yeong Zee Kin SAR stated in *Robin Duane Littau v Astrata (Asia Pacific) Pte Ltd* [2011] SGHC 61 as follows:

> 19 On 13 December 2010, as parties had reached an impasse on the list of keywords, I fixed the application for a special hearing date in order to determine relevance of the disputed keywords. *To facilitate parties' submissions and assist the court in determining whether any particular keyword was relevant, the defendant was permitted to run a preliminary search against the forensic images of the seized items using the list of keywords that it had proposed. However, this preliminary search was intended solely for the purpose of identifying the number of hits* – ie instances of documents which corresponded to the keyword – and the defendant was not permitted to view any of the documents forming the search results.
>
> 20 As it turned out, after performing this preliminary search using the defendant's proposed list of 251 keywords, parties were able to agree to abandon keywords which returned no hits — and the defendant also withdrew two additional keywords — leaving only 92 disputed keywords on which a ruling was required. During submissions on 12 January 2011, the preliminary [search] also provided much assistance, particularly when refining the disputed keywords and determining the search conditions which ought to be attached to the keywords when the search is eventually performed. In the present case, the preliminary search was conducted at a cost of only $500 to the defendant, who had volunteered to bear this cost. Given its relatively low cost and usefulness during arguments before me, *I think that this practice of performing a preliminary search using disputed keywords in order to identify the number of hits should, so far as practicable, be adopted in all cases where keywords are disputed. This will help to identify red herrings (ie keywords which yield no hits) and assist parties to refine search conditions or the keywords proper, whether as part of negotiations or during arguments before the court.*
>
> [emphasis added]

47 Similarly, Senior Master Whitaker observed as follows in *Goodale* (cited at [33] above):

> 24 The next question that arises is: is it going to be a question of simply running the 31 suggested 'key words' across his system and across those of the other three? We do not know at the moment what the likely result of that is going to be, because we do not know enough yet about Palmer, Bradshaw and Piper as to whether they kept anything on their systems and on their shared drives.
>
> 25 *It seems to me that the proper way of going about this in respect of the individuals is that the limited searches that could be run on the MEDS system should be run but without the actual physical review and production of those documents in the first place. We need to know how many documents each of the 31 terms are going to turn up to establish whether any may require a degree of fine-tuning. We also need to know the total number of documents that respond to this collection of search terms (acknowledging that the same documents may respond to more than one search term in many cases).* There is probably going to be some question over whether all the 31 key words are necessary, because what little bit of sampling that has been done seems to reveal, that quite a few of them produce nil returns anyway. We shall see whether that is the case in respect of Palmer, Bradshaw and Piper as well as Marteau when those searches are run.
>
> 26 That searching, because it is going to be done in a comparatively simple way, without using specialist software at this stage, is just going to give us the potential numbers of documents. Similarly, doing the same type of search in respect of the MEDS system for the 31 terms but only

in respect of each of the key witnesses, will give us the potential number of documents in respect of that as well. *It is at that stage, when that crude way of finding out what documents might be in existence is completed, that a service provider will have to be agreed between the parties, and will have to be instructed to look at what the next stage of the exercise should involve and how much it is going to cost, in order to produce a corpus of documents which is reviewable by both parties.*

27 At the moment we are just staring into open space as to what the volume of the documents produced by a search is going to be. I suspect that in the long run this crude search will not throw up more than a few hundred thousand documents. If that is the case, then this is a prime candidate for the application of software that providers now have, which can de-duplicate that material and render it down to a more sensible size and search it by computer to produce a manageable corpus for human review – which is of course the most expensive part of the exercise. *Indeed, when it comes to review, I am aware of software that will effectively score each document as to its likely relevance and which will enable a prioritisation of categories within the entire document set.*

28 It is also possibly going to be necessary to look at whether we should in fact be running all of these key words at this stage. In my judgment, that is what the exercise should be. There should be disclosure of electronically stored information. It is clear that documents created by these four witnesses exist which are likely to support the claimants' case and damage the defendant's. The only question is how we go about finding them. *I think the proper thing to do is to start with a fairly crude search and then, if the numbers are within reason, to work with experts to render the corpus of documents down and de-duplicate them and then move on to the review stage.* ...

[emphasis added]

48 I have also made the following observations in *Breezeway* (cited at [31] above):

26 In my view, it would be helpful to conceptualise the process of identifying relevant material through keyword searches as an *iterative sieving process*. This coheres with para 43B of the e-discovery PD, which contemplates the conduct of general discovery in stages ... Under this iterative sieving process, the court and the parties endeavour to select the best possible keywords that would avoid sieving out relevant material whilst simultaneously ensuring a practical and workable manner of processing the material at hand. Parties would thereafter clarify and/or narrow search terms as necessary with a collaborative spirit and in good faith, resorting to applications to court only when parties require an arbiter to break the impasse. The court will eventually sanction a final set of search criteria for the purposes of e-discovery ("court-sanctioned search"). [emphasis in original]

49 I need only end by commenting that the iterative method is encapsulated in and is an integral part of Part V of the Supreme Court Practice Directions (2013 Revised Edition) ("the e-discovery PD"). Paragraph 45 exhorts parties to engage in good faith discussions during general discovery, and in so doing, to discuss "whether preliminary searches and/or data sampling are to be conducted and the giving of discovery in stages according to an agreed schedule". This is elaborated in the template e-discovery plan at paragraph 1 of Appendix E Part 2 in the following terms:

(e) **Preliminary search.** A preliminary search of the repositories identified in sub-paragraph (d) above is to be conducted forthwith. Such preliminary search is limited to providing information relating to the number of hits and/or the number of documents containing the keywords. Parties

shall review the search results within two (2) days of being provided with the same; and within a further five (5) days, parties shall meet to discuss whether the keywords and/or the repositories identified in sub-paragraph (d) above need to be revised. Parties agree to abandon any keywords with no hits and to review any keywords with hits exceeding [*insert a figure*, eg *10,000*] for the purpose of constraining the keywords. Unless mutually agreed, no new keywords may be introduced following the performance of the preliminary search.

(f) **Data sampling.** Parties agree to perform a reasonable search of the following repositories in sub-paragraph (d) above: [*insert a sample of the custodians and repositories by referencing the table in sub-paragraph (d)*]. Parties shall review the search results within seven (7) days of being provided with the same; and within a further seven (7) days, parties shall meet to discuss whether the keywords and/or the repositories identified in sub-paragraph (d) above need to be revised. Data sampling in accordance with the terms of this sub-paragraph shall be performed no more than twice.

[emphasis in bold and italics in original]

Summary of principles

50 Having set out the relevant case law and commentary on e-discovery generally and keyword searches in particular, I will now summarise my views on these issues in the hope of providing some degree of guidance to parties and/or their lawyers in the future.

51 In keyword searches, there will always remain a risk that relevant documents may not be identified regardless of the number and/or type of keywords used. Broadly speaking, relevant documents which are not identified can be termed "false negatives". However, it must be remembered that even where traditional methods of discovery are utilised (*ie*, by visual inspection of documents for relevance), there is also a risk of false negatives due to human error. Whether this latter risk is lower or higher than the risk of false negatives in keyword searches will depend considerably upon the type of search technology being used in the keyword searches as well as the particular keywords which are used in the searches.

52 The converse type of errors in keyword searches can, broadly speaking, be termed as "false positives". This denotes *irrelevant* documents which are identified by keyword searches.

53 Given that any keyword or set of keywords will produce false positives and false negatives, the concept of relevance (as traditionally understood) does not apply directly to the issue of which keywords should be used in searches of the relevant electronic devices. In other words, it is not fruitful to focus on the concept of relevance in relation to keywords. It must be remembered that the use of keyword searches can come in at a stage earlier than the traditional discovery process, *eg*, the use of keywords by a litigant to identify documents which are then reviewed by their solicitors for the purpose of giving general discovery. What a keyword search does is to identify documents containing the keywords – the search creates a subset of documents from the universal set containing all the documents in the relevant electronic devices. Within this subset there will exist relevant and irrelevant documents, just as there will exist relevant and irrelevant documents outside the subset. The best keyword searches are those that maximise the number of relevant documents and minimise the number of irrelevant documents within the subset. Such searches would be said to have high *accuracy*, *ie*, they have a low proportion of false negatives and of false positives to the total number of documents within the universal set. It is this concept of accuracy which is more pertinent in this context than the concept of relevance.

54 Thus, whenever the question arises as to whether a particular keyword or set of keywords should be used, one of the salient questions is whether it has high accuracy. All other things being equal, the higher the accuracy of the proposed keyword or set of keywords, the more likely it will be that the court will grant an order that those keywords be used. The concept of accuracy, in turn, has two facets. The inclusionary facet of accuracy deals with the correspondence between, on the one hand, the keywords, and, on the other hand, the documents sought and the issues in dispute. Generally speaking, the greater the correspondence between the keywords and the issues in dispute, the more likely it is that the keywords will result in relevant documents being included in the subset (and the lower the proportion of false negatives in the subset to the total number of documents in the universal set). Conversely, the exclusionary facet of accuracy deals with the exclusion of irrelevant documents from the subset. Generally speaking, where the keywords are not common words, it is more likely that irrelevant documents will be excluded from the subset (and the lower the proportion of false positives in the subset to the total number of documents in the universal set).

55 How should we approach the selection of keywords in order to increase the level of accuracy? In *Breezeway Overseas Ltd v UBS AG* [2012] SGHC 41 Yeong Zee Kin SAR had recommended the following approach:

> 28 …*First*, commence with the specific before expansion to broader search terms. Specific search terms would include the following:
>
> (a) **Unique reference numbers.** For example, bank account numbers or client account numbers where the context is in a banking relationship. In the context of other commercial transactions, if one party has in place a file reference number or account identification number, these may be used as well. This very closely approximates the traditional paper filing system.
>
> (b) **Names of specific projects.** This can be an important keyword particularly where the dispute arises from a developmental project or commercial transaction which has been assigned a project name, eg *Robin Duane Littau v Astrata (Asia Pacific) Pte Ltd*.
>
> (c) **Keywords which identify the key witnesses (or custodians).** For example, e-mail addresses, contact numbers and names or initials. Search terms may be formulated based on such keywords. For example, e-mail addresses of two key witnesses appearing in the same e-mail in order to identify e-mail conversations between them, eg *Robin Duane Littau v Astrata (Asia Pacific) Pte Ltd*.
>
> (d) **Significant events and locations.** Depending on the facts, there may have been a significant meeting which took place. It has proven useful in some cases to make use of the meeting location or a short-hand reference to a key meeting as a search term. This may identify correspondence and documents that have been generated surrounding the event. However, care must be taken in selection of locations. In the present case, there were some meetings which took place at the offices of UBS. The search term "Suntec meeting" did not turn up any hits in the preliminary search. Using "Suntec" alone would have turned up too many, as the standard e-mail footer from employees of UBS contains their office address, which is located in Suntec. However, locations have been useful as a keyword for some other cases, eg meeting at the lobby of Furama hotel.
>
> 29 *Next*, we can consider search terms incorporating keywords which are unique to the facts of the case or the context of the dispute. The capabilities of the search engine – viz the search operators available – to be used are particularly important for this stage.

(a) **Product names.** Where the dispute centres around the purchase of certain product, the product may be used. However, one should bear in mind the facts of the case. Product names alone would, for example, be unsuitable if the product is one commonly sold by one of the parties.

(b) **Unique phrases.** For some industries, there may be unique terms or phrases that can be used to identify significant correspondence.

30 As has been observed elsewhere, we should avoid words which are commonly used, either in daily usage or in the context of the industry: see *Robin Duane Littau v Astrata (Asia Pacific) Pte Ltd*. Additionally, words which are legal concepts that may be part of the lawyers' vocabulary and thought process may not always be helpful as keywords, eg breach of contract, confidential, etc. If such keywords are to be used, the search term will have to be carefully crafted.

3 1 ***The choice of keywords and formulation of search terms depends on the facts of each case and the issues in dispute. The guidelines and approach set out above provides a means to navigate these waters, but cannot be in any way considered definitive or exhaustive*** **.** Further, familiarity with the capabilities of the search engine to be used will in almost all cases be particularly helpful. This will assist the court in formulating search terms which could include keywords which are common words, as in this case (see below).

[emphasis in bold and italics in original, emphasis in bold italics added]

56 The recommended approach increases the level of accuracy but the keywords that ought to be used ultimately turns on the facts and issues of each case. Equally important to the selection of appropriate keywords is the necessity for counsel to be familiar with the capabilities of the search engine that is used in order to increase the level of accuracy of the search results. For example, if the keyword search technology being used is able to utilise Boolean search operators such as "not", the exclusionary quality of the keyword or set of keywords will accordingly be improved.

57 The other factor which should also be considered by the courts pertains to the size of the subset which is produced by the relevant search, *ie*, the number of hits which are produced in proportion to the universal set. If the subset is extremely large (by reference to, *inter alia*, the value and/or importance of the underlying dispute, and the importance of the documents sought to the issues in the underlying dispute), this will generally count as a factor weighing against an order for discovery by way of those keywords. This is a nod to efficiency and economy in civil procedure. In such circumstances, it would usually be prudent for parties to attempt to reach an agreement to modify the keywords so as to reduce the size of the subset. In this connection, it would be helpful if preliminary searches using the proposed keywords were conducted prior to the negotiations between the parties and/or an application to court for discovery, so that there would be some factual basis for the negotiations and/or application.

58 Where a dispute arises as to the keywords which are to be used in a search, the court will have to carefully balance these two factors, *viz*, accuracy and the size of the subset, as well as other proportionality and economy considerations such as the importance of the documents sought through these keywords to the issues in dispute, the value and importance of the dispute, and the relative financial resources of the parties to the discovery application.

59 It would be desirable for parties to propose keywords and to cooperate to arrive at an acceptable compromise by way of an agreed list of keywords. Such cooperation would be desirable because it reduces the cost of litigation for clients. *Paul & Baron* have suggested that it may in fact

be in both parties' interest for such collaboration and cooperation to occur:

> [27] Quite simply, as courts and commentators have increasingly come to expressly recognize, the volume and complexity of electronically stored information demand new forms of collaboration. In turn, in many such instances, a tipping point can be said to have been reached where the game theoretical aspects of litigation practice, dictating what is in one's self-interest, have necessarily changed. Without greater cooperation among adversaries, parties are doomed to any number of defeating consequences, not the least of which will be a real or perceived information "gap" in ferreting out evidence.

60 The benefit of collaboration has not escaped legislative nor judicial notice. This is neatly summarised in the following passage from *Fermin Aldabe v Standard Chartered Bank* [2009] SGHC 194:

> 35 … In Practice Direction 3 of 2009, parties are "encouraged to collaborate in good faith and agree on issues relating to the discovery and inspection of electronically stored documents"; … This is an approach which mirrors the approach taken in both the United States and United Kingdom. Under paragraph 2A.2 of the Practice Direction to Part 31 of the UK Civil Procedural Rules, parties are required to discuss any issues that may arise regarding searches for and the preservation of electronic documents before the first Case Management Conference. … A similar process is set out in Rule 26(f) of the US Federal Rules of Civil Procedure. An obligation is placed on parties to confer, 21 days before a scheduling conference, to discuss any issues relating to preserving discoverable information and to develop a discovery plan that addresses, … Indeed, in the recent decision in *Digicel (St. Lucia) Ltd & Ors v Cable & Wireless Plc & Ors* [2008] EWHC 2522 (Ch), at [47], the court highlighted the potential pitfalls where parties fail to meet for discussion:
>
> > This case provides an opportunity for the Court to emphasise something mentioned in Part 31 Practice Direction which the parties in the present case disregarded. Paragraph 2A.2 of the Practice Direction states that the parties should at an early stage in the litigation discuss issues that may arise regarding searches for electronic documents. Paragraph 2A.5 of the PD states that where key word searches are used they should be agreed as far as possible between the parties. Neither side paid attention to this advice. In this application the focus is upon the steps taken by the Defendants. They did not discuss the issues that might arise regarding searches for electronic documents and they used key word searches which they had not agreed in advance or attempted to agree in advance with the Claimants. *The result is that the unilateral decisions made by the Defendants' solicitors are now under challenge* and need to be scrutinised by the Court. If the Court takes the view that the Defendants' solicitors' key word searches were inadequate when they were first carried out and that a wider search should have been carried out, the Defendants' solicitors' unilateral action has exposed the Defendants to the risk that the Court may require the exercise of searching to be done a second time, *with the overall cost of two searches being significantly higher than the cost of a wider search carried out on the first occasion*. [Emphasis in original]

61 During the discussions, both parties should bear in mind that the process of keyword searching is best carried out through an *iterative* process. Thus, what would generally be sensible for the party from whom discovery is sought is to conduct preliminary searches on the material electronic devices to show how many hits are produced by the keywords which he proposed or which the party seeking discovery proposed. These proposed keywords, together with the number of hits produced by the preliminary searches, would assist in further discussions between the parties if further modification of

the list of keywords is required. The conduct of preliminary searches would also usually save considerable time, money and effort which might otherwise be spent on arguments in the abstract on whether a particular keyword or set of keywords was over-inclusive or under-inclusive.

62 In addition, the party from whom discovery is sought should generally bear in mind the fact that e-discovery by way of keyword searches is intended to relieve him, to some extent, of the burden of reviewing all documents in his possession, custody or power for relevance and/or privilege. As I stated in *Breezeway* (at [33]):

> 33 ... At bottom, the e-discovery PD is designed to keep costs proportionate, relieving the party giving discovery of the need to conduct costly and time-consuming ocular review ...

The purpose of e-discovery by way of keyword searches is not to disadvantage the party giving discovery, but rather to put him at an advantage in terms of the cost of complying with his discovery obligations *vis-à-vis* all documents in his possession, custody or power.

63 Having said that, however, where disputes about particular keywords or about the entire list of proposed keywords arise, the courts should, in resolving such disputes, endeavour to aid the party seeking discovery by giving more weight to his proposed keywords in deciding which keywords should be included in the order for discovery. This is because if the party from whom discovery is sought complies with the court order for discovery by way of particular keywords, that will discharge his discovery obligations at that stage. In *Sanae Achar* (cited at [32] above), I stated as follows:

> 23 Having explained the basis for ordering discovery of the Category 1, 2, and 3 Documents, I briefly turn to the extent of Achar's and Sci-Gen's obligations. Pursuant to my discovery order, Achar must, *inter alia*, disclose the documents specified in the order, carry out a search to the extent stated in the order, and disclose any documents located as a result of that search. *So long as Achar has complied with the terms of that order, as well as all the necessary requirements stated in the Rules of Court*, ***Sci-Gen would have to accept that Achar had fulfilled her discovery obligations, notwithstanding the fact that there could well be e-mails not caught by the search engine employed*** . As Morgan J (echoing a point made by Jacob LJ in *Nichia Corporation v Argos Limited* [2007] EWCA Civ 741 at [50] to [52]) articulated in *Digicel (St Lucia) Ltd v Cable & Wireless Plc* [2008] EWHC 2522 (Ch) at [46]:
>
> > … [T]he [discovery] rules do not require that no stone should be left unturned. This may mean that a relevant document, even 'a smoking gun' is not found. This attitude is justified by considerations of proportionality. …
>
> In this regard, it would be best if the parties can, prior to any search, agree on which search engine or software is to be used, the preparation of the search engine prior to conducting the searches (*eg*, updating the search index or causing a fresh search index to be made) and how searches are to be conducted. This would minimise potential disputes as to whether the parties have discharged their discovery obligations.
>
> [emphasis added in italics and bold italics]

64 In *Breezeway* (cited at [31] above), I reiterated this point:

> 30 The upshot of the above observations is that the concept of "*prima facie* relevance" refers to the notion that the party giving discovery is not *required* to review the search results of the court-sanctioned search for relevance. The search results of the court-sanctioned search are

"*prima facie* relevant" in the sense that the party giving discovery will be deemed to have complied with his obligation to provide all relevant documents under the general discovery process (see O 24 r 1 of the Rules of Court).

31 In this regard, one should be keenly aware of the conceptual distinction between the *obligation to give discovery* and the concept of *relevance* in the context of discovery. Although both concepts are related because the party giving discovery has the *obligation to give discovery* of all *relevant* material, ***the e-discovery PD makes it possible for that party to fulfil his discovery obligations by giving discovery of the results of a court-sanctioned search, regardless of whether such search results are over- or under-inclusive vis-à-vis the identification of relevant material.***

32 With regard to the *obligation to give discovery*, ***as long as the party giving discovery complies with the terms of the court-sanctioned search, as well as with all the necessary requirements as stated in the Rules of Court, the party entitled to discovery "would have to accept that [the party giving discovery] had fulfilled [his or] her discovery obligations, notwithstanding the fact that there could well be [documents] not caught by the search engine employed"*** ...

[emphasis in original in italics; emphasis added in bold italics]

65 Where the keywords in issue are those which are proposed by the party seeking discovery, it will generally be unnecessary to be concerned about false negatives. In such circumstances, it will generally be reasonable to assume that that party would have carefully considered the risk of false negatives and tailored the keywords which he proposes to minimise this risk. Thus, in general the only concerns will be in relation to the probability of false positives and the likely number of false positives.

66 Once the keyword search which is agreed by the parties or which is ordered by the court is carried out, the party giving discovery may still conduct a post-search review of the hits to sieve out irrelevant documents. However, this review must at present be done in the traditional manner. In *Breezeway* (cited at [31] above), I observed as follows:

33 However, the fact that the obligation to give discovery is fulfilled by the party giving discovery of the results of a court-sanctioned search does not mean that the results of the search are deemed relevant in the sense that the party giving discovery is not entitled to conduct post court-sanctioned search reviews. At bottom, the e-discovery PD is designed to keep costs proportionate, relieving the party giving discovery of the need to conduct costly and time-consuming ocular review of all the documents in his possession, custody or power (see [22] above). *The e-discovery PD was not intended to prevent the party giving discovery from undertaking a post court-sanctioned search review to remove documents that are irrelevant to the issues in dispute. But any such further review would be outside the ambit of the e-discovery PD and the decision to remove any document on the ground of irrelevance must be done by way of ocular review. This means that every document a party removes in a post court-sanctioned search review on the basis that it is irrelevant must be processed in the traditional manner, ie, manually examined and subsequently considered irrelevant by a solicitor familiar with the issues in dispute (or the party, in the case of a litigant in person). As this process is usually an expense unreasonably incurred the party electing to do this will not generally be entitled to recover the costs of the post court-sanctioned search review in the event that costs are eventually awarded in his favour.*

[emphasis added]

67 The party giving discovery may also wish to withhold certain documents on the basis of privilege, or to redact irrelevant confidential information which is contained in discloseable documents. He may do so provided the review is done in the traditional manner. In *Breezeway* (cited at [31] above), I observed as follows:

> 34 Other than a post court-sanctioned search review for relevance, it is clear that the party *giving* discovery may conduct a post court-sanctioned search review for *privileged and/or confidential material* ... Again, the same conditions apply as in a post court-sanctioned search review for relevance, *ie*, any documents may only be withheld from discovery on the ground of privilege and/or confidentiality pursuant to ocular review, and costs for this extra step will not generally be allowed. [emphasis in original]

68 The recognition of e-discovery by way of the e-discovery PD represents an attempt to reach a pragmatic compromise between "[t]he perennial tension in the law of civil procedure, *viz*, the attempt to achieve both justice and efficiency" (*Breezeway* at [20], quoted at [31] above) using modern technology. Applied in a pragmatic and collaborative way, it is hoped that e-discovery will help to ensure that the cost of litigation is not disproportionate, having in mind the importance of the documents sought to the issues in the underlying dispute, the value of the underlying dispute between the parties, and the relative financial resources of the parties.

69 Before I explain why I dismissed the appeals on the facts, I pause here to set out some useful observations in *Disclosure* (at para 7.12) with which I agree:

> In some circumstances, it may be reasonable to search for electronic documents by means of key word searches or other automated methods of searching (agreed as far as possible between the parties) if a full review of each document and every document would be unreasonable. *However, it is often insufficient to use simple key word searches or other automated methods alone. An inappropriate use of automated searches may result in important documents not being found or excessive irrelevant documents are found. Hence automated searches will usually need to be supplemented by additional techniques such as individually reviewing certain documents or categories of documents (for example, important documents generated by key personnel).* [emphasis added]

This serves as a useful reminder that it may sometimes be reasonable for keyword searches to be supplemented or complemented by the traditional review of particular subsets or classes of documents or documents generated by particular persons. An example will be where the keywords are selected unilaterally by one party prior to general discovery, without the benefit of a collaborative discussion with his adversary: the focus here remains on the party providing relevant documents since he is working in a silo. Where keywords are selected by agreement after a collaborative discussion (whether or not a formal electronic discovery plan is drawn up) or ordered by the court pursuant to an application by summons, then the focus shifts from giving discovery of relevant documents to the fulfilment of the discovery obligation by conducting the search and producing its results. Having said that, this dichotomy of approaches may need to be recalibrated as the state of keyword search technology improves as well as the availability, reliability and cost of alternative search technologies.

Application of the law to the facts

70 Having set out the applicable legal principles above, I now explain why I dismissed the appeals subject to the variations mentioned at [23] above.

71 I was of the opinion that there was little basis for Promedia's and Streetdirectory's objection that the keywords proposed by GYP were over-broad in that they would result in a large number of hits *per se*, given that Promedia and Streetdirectory had not conducted a preliminary search on the relevant electronic devices to ascertain precisely how many hits were produced by the proposed keywords. While the concerns of Promedia and Streetdirectory may turn out to be well-founded, it was also possible that only a reasonable or proportionate number of hits would be produced by the proposed keywords. In addition, the size of the subset which was likely to be produced by the search was only one of the factors which the court had to consider. The other pertinent factor was that of the accuracy of the keywords which were in issue.

72 I was of the view that searches conducted using the seven keywords which were common to both appeals (see [22] above) would have a reasonable degree of inclusionary accuracy, *ie*, there would be a reasonably low likelihood of false negatives in the hits produced by the search. While some of the keywords, *eg*, "copy", "delete" and "directories", were words which were likely to have been commonly used in irrelevant contexts particularly when the nature of Promedia's and Streetdirectory's businesses is taken into account, I was of the view that the issue of whether the keywords had a low degree of exclusionary accuracy (*ie*, whether there was a significant risk of too many false positives in the hits produced by the search) was mitigated significantly in the particular circumstances of these appeals. This was because pursuant to the AR's orders in SUM 773/2012 and SUM 774/2012 (see, for Suit 913, pp 3 and 4 of the draft order annexed to the notice of appeal in RA 421/2012, and, for Suit 914, pp 4 and 5 of the annex to Order of Court No 5977 of 2012), all parties would not review documents which were produced by searches with more than 5,000 hits. No appeal was filed against this part of the AR's decision. In other words, the maximum number of false positives which would be produced by the keywords and which would be discloseable (subject to further manual review for relevance and/or privilege) would be 5,000 documents.

73 One of the arguments which Promedia and Streetdirectory raised was that the terms "seed" and "seeds", used in the sense described in GYP's pleadings (see [12(b)] above), were unique to GYP. GYP accepted that it had coined the terms "seed" and "seeds" in this context. In other words, the terms "seed" and "seeds" were used in a unique sense by GYP and were not generally used in this sense by persons in this particular industry. This indicated that searches using these keywords would have a low degree of inclusionary accuracy, for the simple reason that Promedia and/or Streetdirectory (or their employees) would generally be unlikely to describe the Seeds as such. Nonetheless, I dismissed the appeals in respect of the terms "seed" and "seeds" because Promedia and Streetdirectory were aware of the use of these terms in this context after GYP filed an affidavit by Freddie Tan Poh Chye dated 17 September 2010 in Suit 914 ("Freddie Tan's Affidavit"), para 12 of which referred to the Seeds as defined in GYP's pleadings. Thus, there was a possibility which could not be discounted that Promedia and/or Streetdirectory had used the terms "seed" or "seeds" in their internal correspondence and/or documents after that date. Furthermore, Promedia had in fact removed the Seeds after Freddie Tan's Affidavit was filed: see para 62 of Teo Chai Tiam's affidavit dated 27 March 2012. In my view, it would be wrong in principle to exclude these terms from the search given that there was a not insignificant possibility that they may produce relevant documents. On the facts of this case, I was not persuaded that there was sufficient reason to depart from the *prima facie* position that greater weight should be given to the keywords proposed by the party seeking disclosure (see [63]–[64] above), particularly because the maximum number of hits which were discloseable was 5,000 (see [72] above).

Conclusion

74 For the reasons above, I dismissed the appeals subject to the variations mentioned at [23] above.

[note: 1] statement of claim ("SOC") (amendment no 3) in Suit 913, para 4

[note: 2] SOC (amendment no 3) in Suit 913, para 12

[note: 3] SOC (amendment no 3) in Suit 914, para 12

[note: 4] SOC (amendment no 3) in Suit 913, para 13; SOC (amendment no 3) in Suit 914, para 13

[note: 5] SOC (amendment no 3) in Suit 913, para 20; SOC (amendment no 3) in Suit 914, para 18

[note: 6] defence and counterclaim (D&CC) (amendment no 3) in Suit 913, para 5

[note: 7] D&CC (amendment no 3) in Suit 913, para 13

[note: 8] D&CC (amendment no 3) in Suit 913, para 13(ii)

[note: 9] D&CC (amendment no 3) in Suit 913, para 13(iii)

[note: 10] D&CC (amendment no 3) in Suit 913, para 22

[note: 11] D&CC (amendment no 3) in Suit 913, para 29

[note: 12] D&CC (amendment no 3) in Suit 913, paras 30, 31 and 32

[note: 13] D&CC (amendment no 3) in Suit 913, para 44

[note: 14] Reply and Defence to Counterclaim (amendment no 3) in Suit 913, paras 21 to 24

[note: 15] Defence (amendment no 2) in Suit 914, para 16

[note: 16] Defence (amendment no 2) in Suit 914, para 19

[note: 17] Accessed at http://jolt.richmond.edu/v13i3/article10.pdf on 24 April 2013

Copyright © Government of Singapore.