From Information to Knowledge
M.A. Biasiotti and S. Faro (Eds.)
IOS Press, 2011

doi:10.3233/978-1-60750-988-2-83

Proper Treatment of Gaps in Legal Data and Commercial Electronic Legal Search Services

Melih KARAKULLUKÇU [a], Yeşim ERIBOL [b] and Tuğba KARAER [c]
[a] *Founding Partner, Karakullukçu Danışmanlık Ltd. Şti. (Karakullukcu Consulting LLP), owner of the Turkish electronic legal search service Kanunum*
[b] *Legal Consultant and Kanunum domain expert*
[c] *Kanunum domain expert; post-graduate student, Istanbul University Law Department*

Abstract. The paper presents the technical and legal issues that an electronic legal search services has to face while dealing with the problem of missing, unreliable or quasi-official legal data. Technically the problem requires the sifting of acquired legal texts in terms of importance and reliability, and coming up with user friendly matrices and expressions to present information to the researcher that conveys texts with their accompanying degrees of ambiguities, their varying reliability and their connections to other texts in the "corpus." Ultimately the aim is to formulate and delineate (for the user) the limits of reliability of any given legal electronic information. Moreover the initiatives carried on by the authors' on line legal service (Kanunum) to make more legal texts available electronically in Turkey from the official hand that produced these texts and actions are presented.

Keywords. Online access to legal information, online legal search services, case law.

Introduction

Users approach private and public legal electronic search tools to find answers to their legal questions. Meeting the users' objective involves a formidable research operation for these services. This is because proper answers in the legal domain require a comprehensive search through the large, nebulous, and hermeneutically connected corpus we call "the law." "The law" consists of not just statutes, regulations and high court decisions but recorded regulatory actions, lower court rulings, administrative rulings, official answers to citizens' queries and the like. In most jurisdictions it is the administrative state's practices with regard to a statute or court decision that ultimately determines what "the law" is for a citizen. Further, it is only when analyzed in tandem with each other that any of this can be properly understood. Given the size of most EU states, with their numerous administrative units, this means that part of the corpus we call "the law" will always be missing from electronic (and even print) publication. Yet, because of the coherence of the legal domain, the electronic legal search service will never be able to ignore the missing data without consideration. What follows is an overview of how we at Kanunum view and tackle the problem of missing information. But first a word on what "Kanunum" is and what we mean by "information" or "data."

Kanunum was launched by a Turkish SME in 2008 as a legal informatics project intended to deliver internet search tools for Turkish professionals. In late 2009, it went live on www.kanunum.com as an online, up-to-date and searchable electronic compilation of current statutes and their complete textual histories, with a collection of high court decisions. In mid-2011 Kanunum became a for-fee service with an expanded database, but it remains dependent on private funding and R&D state grants. It is run by a small staff consisting of two computer scientists specializing in Natural Language Processing technologies, a lead engineer, several developers and legal editors, and an Editorial Board of five lawyers (domain experts). On the R&D side we seek to further reduce human involvement in the expansion and daily maintenance of legal data and improve our search performance. As for our general direction, we have spent some time defining our role as a company: Kanunum is a publisher of legal material in an online electronic environment. Thus we see modern online legal search services as a continuation of print legal publishing, sharing its goals of presenting legal texts in an accurate, easy-to-understand, and useful manner. These being the goals, we have come to define our primary opportunity and challenge in the electronic online environment as one of conversion: we aim to take the unorganized and changing mass of legal electronic material on and off the web and convert it into an online service that produces user friendly screens of correct, time-sensitive, and relevant legal texts and contexts in response to user queries.

In light of that perspective, by "data" we mean material that enables us to present an official normative text with its correct wording and essential context at a given point in time (e.g. the wording of a statute section on a given day, with references to texts such as interpreting regulations that affect the section's reading on that day). With that in mind, we simply and anecdotally detail below why important "data" will always be incomplete, why that is a problem and what we do when data is missing or unreliable. Second, we discuss the legal initiative we have pursued in trying to get authorities to make more "data" available.

1. Managing Information Gaps as Electronic Legal Publishing

As most other online legal search services, Kanunum's users come to the webiste with specific queries which they subsequently type in a search box. The service produces a list of results with snippets where search terms are highlighted. The user clicks on the result that seems pertinent and the requested text is displayed. In the specific case of statutes, the content viewer displays a timeline at the top indicating textual amendments, while separate sections on the screen show a selection of other legislative developments and judicial decisions that have had the effect of a textual amendment for the statute being viewed. This information is repeated in modified particularized form for all the statute sections shown on the screen. Thus, when looking at a statute section, one can see the dates on which the particular section was amended, temporarily suspended or struck by the high court, or repealed. When one clicks on a marked point on the timeline, one sees in a new window the legislative or judicial text affecting the change for that date. Apart from such textual information, the content viewer of the statute tries to show appeals court decisions or regulations that interpret the statute and/or the particular section. The content viewer also tries to warn the user of amendments or textual changes on a section that will or that might come into effect at a future date.

To produce this service, Kanunum depends first on Natural Language Processing technologies to search through its database. As for the database of legal material, the texts are gathered primarily from official electronic sources. To achieve text-normalization, remain current with the daily amendments and maintain a connected database, we rely on a workflow combining software tools developed in-house and editors trained within the project.

As we gather, compile, search through and display legal texts, the process produces consequential gaps and mistakes in information in a number of ways. The first place where information contamination might occur in our work is the technical leg of the workflow, that is the software applications that the editors use to enter the results of their daily textual analyses of the official gazette and judicial reporters, and the data management system which is used to maintain and display texts and relational data. However much an institution may invest in strengthening these software tools, the unpredictable software mishaps (termed "bugs" by developers) are unavoidable, especially during the earlier phases of development. In the electronic legal publishing context, the significance of these mishaps can greatly vary. To provide one example, we have discovered during the trial phases of a tool that automatically tagged XML files where it found references to other legal texts, that the tool surreptitiously (and seemingly randomly) deleted original textual material while tagging the XML files. In parts, the deletions were extensive, rendering the texts incomprehensible. Often, however, they were minor enough that the texts read fine and the gaps were not apparent, rendering the technical error difficult to detect early on.

The other place where errors cause gaps or mistakes in information is where editors engage in textual analysis, modifying texts and creating relational data and metadata. Common editorial errors we encounter when checking for mistakes include entering a wrong effective date for an amendment, creating a relation between a high court decision and legislative text suggesting judicial review where none exists, or entering the wrong official gazette date.

Finally, in addition to such inadvertent intrusions, gaps in data can also be the result of the publication decisions of the various branches of government. As mentioned, for a user who during her search encounters a legislative text, the proper interpretation of the legislation would require a knowledge of related regulations, regulatory actions, and administrative practices. Yet, not all regulations are officially published, regulatory actions are not always reported in official reporters, and many court decisions that strike or suspend a regulatory section may not be available in any official source. Furthermore, in the online electronic environment texts are displayed for the user upon demand and there is a presumption on the users' side that the information presented to them is current as of the time the material is being viewed. So, even where authorities may have the willingness to share information, they may lack the ability to meet the heightened expectations of online users.

These problems we believe are bound to exist to varying degrees in any product and jurisdiction. Especially where resources are limited, completely eradicating the problem of missing or erroneous data is not an option. On the other hand, as indicated, certain gaps can be so invidious so as to render electronic legal research services unusable. Given our resources at Kanunum, we tackle the problem of missing information through two methods primarily – by screening data for errors and by making sure that we appropriately warn users against inaccuracies.

As for the former, we selectively screen for errors and gaps in our data. For that we have had to make unattractive determinations that certain gaps in data are less

important than others. To remain consistent in those decisions, we have agreed on two factors when determining the significance of gaps: (i) obviousness and (ii) potential harm to the user. Thus, in the technical mishap described above, we find those deletions of texts that are extensive and obvious to be less significant than small deletions which change meaning without rendering the text incomprehensible. Or, to give another example, we feel that providing the wrong official gazette number when citing a statute in an official court document is less harmful for a user than relying on statutory language that is not yet in effect. Thus, when creating automatic and manual checks for errors, we prioritize certain types of information and relations.

Second, we believe users should be informed about the limits of the accuracy and currency of the information that they see on the screens of private electronic legal publishers. While providing global warnings that official sources should always be consulted, publishers should also include specific directions where possible. For instance, when displaying the textual histories of statutes we take care to convey that the history includes direct textual amendments, as opposed to all legislative actions such as budgets that may have temporarily suspended a statutory section during a past budget year.

As for the gaps that emerge when authorities fail to make texts officially available, we still use the above methods in facing such gaps. However, there is of course an additional option when texts are withheld; one can ask for them. In the next section we briefly discuss how we have explored that option in our jurisdiction.

2. Convincing Authorities to Make Case Law Available – A Test Case

As in most other jurisdictions, court decisions and especially high court decisions constitute a significant source of legal authority in Turkey; they are precedent. However, only a selection of cases are published in official reporters and websites, and the official print and electronic publications frequently show excerpts or summaries rather than the verbatim texts of the rulings. Of the hundreds of thousands of high court decisions that are not officially published some important ones surface in private publications and websites. It is not uncommon for private legal electronic search services to boast that their case selection contains important decisions that cannot be found elsewhere. This being the case, a serious – if not the most significant – consideration for a company in the electronic legal publishing business is access to these "unpublished" decisions.

Kanunum had no special access to the "unpublished" decisions. Hence we decided to try a different route. First, we wanted to legally press the point that as a rule the state should electronically publish all court decisions, and especially high court decisions, in their entirety at the time they are entered. We believe decisions should escape official electronic publication only exceptionally and for well-defined reasons, not as a default. Second, we wanted to enter into a public and official discussion on the nature and basis of the manner in which some private parties such as publishers were obtaining otherwise "unpublished" decisions.

With that aim, as early as May 2008, we wrote to *Danıştay*, the Turkish administrative appeals court[1]. In our letter, we explained that we were a company

[1] *Danıştay*, translated as the "Council of State" in some official documents, is the highest court of appeals in administrative cases in Turkey.

intending to electronically publish official legal texts such as court decisions and that we therefore would be grateful if the court provided guidance on how to acquire the administrative appeals decisions that were in digital format. When this initial correspondence did not yield results, the Kanunum domain experts decided to file a formal application with the administrative appeals court under the Turkish Right to Information Act (RIA). In our May 2010 RIA application we requested, (1) the case numbers and dates of all en banc decisions that had been published by the court in electronic or print media, (2) access to the remaining en banc decisions, and (3) information on internal regulations (if any) controlling the practice of publication of court decisions by court judges through private publishers. On 31 May 2010 the administrative appeals court declined all RIA requests, stating that per RIA it could decline the requests either (i) under section 7 of the statute that excepts requests where the institution would have to engage in a "separate or special study, research, or investigation" to produce the demanded information[2] or (ii) under section 27 of the statute that leaves demands for statements of "recommendation and opinion"[3] outside the RIA's scope. The court stated that, furthermore, RIA's ambit would not extend to court activities that were adjudicatory in nature. The court emphasized generally the legal and technical barriers before unlimited access to court decisions. The court referred us to the selection of cases published on their website and to the regulation that defines the procedures for accessing the court's archives.[4] With regard to the latter, the regulatory section cited by the court says, as relevant, that researchers can apply for access to the archives with a statement on the scope of their research.[5]

In June 2010 we petitioned the RIA Board of Review, the administrative body which hears the complaints on declined RIA requests. We explained, among other things, that whether or not RIA excluded from its ambit information requests on adjudicatory activities, the texts of decisions in final judgments were not information pertaining to an adjudicatory activity of the administrative appeals court, and noted that the administrative appeals court maintained these texts as archival material and the legal community relied on them as precedent. Thus the specific request did not fall outside the RIA's ambit in that regard, even assuming that adjudicatory activities were excluded from the RIA. We also noted that our request for information on the procedures controlling the private publication of decisions by court judges was a demand for information on the court's existing rules and practices and not one for a statement of opinion from the court.[6] In July 2010, the RIA Board of Review agreed that the court should provide the information on its rules regulating the practice of private publication of decisions by judges. As for our request for the case numbers and dates of published en banc decisions, the board stated that the court could decline this request under RIA's section 7 'separate research' exception. As for the unpublished decisions, the board deemed the court's reply "adequate."[7] Pursuant to the board decision, the administrative appeals court replied to us in August 2010 that there was

[2] *Bilgi Edinme Hakkı Kanunu*, T.C. Resmi Gazete 25269 (24 October 2003) (The Right to Information Act (RIA)), Section 7.

[3] Id., Section 27.

[4] See the RIA Board of Review decision 2010/1272 (29 July 2010). The RIA Board of Review can be contacted for texts of decisions (http://www.bedk.gov.tr).

[5] *Danıştay Arşiv Yönetmeliği*, T.C. Resmi Gazete 24657 (31 December 2002) (The Regulation on the *Danıştay* Archives).

[6] The authors can be contacted for details of the petitions which are on file with Kanunum (http://www.kanunum.com).

[7] Board of Review decision 2010/1272 (29 July 2010).

no procedure controlling the compilation and publication of court decisions by private publishers. The court's letter added that if we provided the details of the chambers, numbers and dates of the decisions we wanted, the relevant chamber would consider the request.

The Kanunum domain experts decided that Kanunum should file a new RIA request, narrowing down its request so as to escape the section 7 'separate research' exception and to squarely test the RIA Board of Review's stand on the question of the RIA's applicability to information on unpublished court cases. In September 2010 Kanunum filed a new RIA request with the administrative appeals court, this time asking for the case numbers of final decisions entered by one specific chamber on several weeks in June and September 2010. The court declined stating that the demand required "separate research" under section 7, and our petition to the RIA Board of Review was denied on procedural grounds related to the applicable time limits for filing petitions[8]. At this time a Kanunum domain expert, and one of the authors, needed – for her academic research – case law from *Yargıtay* (the Court of Cassation), the highest appeals court in non-administrative cases. The Kanunum domain experts saw this as a new opportunity to obtain a clear decision from the RIA Board of Review, and decided to support and advise the author's private RIA application and anticipated RIA petition.

In February 2011 the author filed a private RIA application with the Court of Cassation and requested the case numbers of one court chamber's final rulings entered between 6-10 December 2010, emphasizing that recent decisions from this chamber were not otherwise available in official publications.[9] The Court of Cassation did not respond, the author successfully petitioned the RIA Board of Review for a response[10], and in their subsequent June 2011 reply the Court of Cassation stated that the courts were outside the ambit of the RIA with regard to their adjudicatory activities and declined the RIA request.[11] The author petitioned the RIA Board of Review in July 2011. The petition stated that the demand for information on the decisions entered upon final judgments were not a request in connection with an adjudicatory activity because the entry of a judgment marked the completion of the adjudicatory process. These decisions were public documents constituting one of the bases of the law. The author also cited the board decision in Kanunum's previous petition as suggestive authority that the request for case law did not fall outside the ambit of the RIA as a request pertaining to an adjudicatory activity. Finally, the petition mentioned that our limited request could neither be denied under the RIA section 7 "separate research" exception.

In August 2011 the RIA Board of Review agreed, 6 votes against 3. The board decided that under the RIA the Court of Cassation had to provide the case numbers pertaining to all the specified decisions. The RIA Board of Review stated that with the entry of judgment, the decisions in these cases ceased to be a part of the adjudicatory process of the Court of Cassation, became archival material and turned into administrative documents belonging to the Turkish people. The board added the decisions were an appropriate subject for an RIA request and the specific request here did not fall under any of the RIA exceptions.[12]

[8] See the RIA Board of Review decision 2010/2159 (30 December 2010).
[9] See the RIA Board of Review decision 2011/1214 (11 August 2011).
[10] See the RIA Board of Review decision 2011/788 (26 May 2011).
[11] See the RIA Board of Review decision 2011/1214 (11 August 2011).
[12] See the RIA Board of Review decision 2011/1214 (11 August 2011).

Of the three dissenting votes, one repeated the view that the decisions were outside the scope of the RIA as adjudicatory documents, citing section 2 of the RIA Regulations[13] interpreting the RIA's intended subject matter. According to this vote, in any event, the RIA request here required "special research" on the part of the Court of Cassation. The dissenting opinion further noted that the idea that under RIA court decisions ceased to be a part of the adjudicatory process upon entry of final judgment, contradicted the RIA's first section which emphasized "democratic and transparent government."[14] The remaining two dissenting votes noted that the RIA was not an appropriate vehicle for demanding information from the judiciary as the act's target was the executive branch. The dissenters suggested that case law should be requested from the courts not through the RIA, but directly, by citing the principle of open and fair trial[15].

3. Conclusion

The completion and responsible management of missing and erroneous texts is a central part of the work of a publisher. We tried to convey through our experience that this is even more so in electronic legal publishing. In connection with our efforts to acquire legal texts from official sources, we suggested that the commercial publisher's struggle has a public good aspect to it. That was part of the motivation in the RIA applications we filed and supported. In that regard, we would like to note that we appreciate the reluctance of the courts and others about expanding the RIA to requests for court documents. On the other hand, we insist that it is not good practice for the state to simultaneously suggest that court decisions are part of "the law" of a country, refuse to officially publish all (or nearly all) of these decisions, and yet make them arbitrarily available to various private publishers.

[13] *Bilgi Edinme Hakkı Kanununun Uygulamasına İlişkin Esas ve Usuller Hakkında Yönetmelik,* Section 2. Resmi Gazete 25445 (27 April 204). The section describes and gives examples of the intended respondents of RIA applications.

[14] The dissenting board member presumably refers to the undesirability of expanding the scope of interference with the activities of the judicial branch, whose independence is a tenet of democratic government.

[15] *Ibidem.*