**105.Legal Case Reports--法律**

1. 数据库网址

https://archive.ics.uci.edu/ml/datasets/Legal+Case+Reports

2. 数据库描述

【1.[数据集名称]数据集由[机构名或人名]采集；】The data used in our experiments were collected by E. Alpaydin, C. Kaynak, from Department of Computer Engineering,Bogazici University at July,1998.【2.用于[什么实验目的]】We used preprocessing programs made available by NIST to extract normalized bitmaps of handwritten digits from a preprinted form.【3】

【4】The database has 5620 samples, respectively belong to optdigits.tra with 3823 samples and optidigits.tes with 1797 samples. The categories of network system include seven categories, as shown in Table 1.

Table 1 Category Distribution of Network System [根据数据库绘制]

|  |  |  |  |
| --- | --- | --- | --- |
| Invasion Categories | optdigits.tra | optdigits.tes | Total Number of Samples |
|  |  |  |  |
|  |  |  |  |
| Total number of samples in total |  |  |  |

|  |  |
| --- | --- |
| **Abstract**: A textual corpus of 4000 legal cases for automatic summarization and citation analysis. For each document we collect catchphrases, citations sentences, citation catchphrases and citation classes. |  |

**Source:**

Filippo Galgani   
galganif **'@'** cse.unsw.edu.au   
School of Computer Science and Engineering   
The Univeristy of New South Wales, Australia

**Data Set Information:**

This dataset contains Australian legal cases from the Federal Court of Australia (FCA). The cases were downloaded from AustLII ([[Web Link]](http://www.austlii.edu.au/)). We included all cases from the year 2006,2007,2008 and 2009. We built it to experiment with automatic summarization and citation analysis. For each document we collected catchphrases, citations sentences, citation catchphrases, and citation classes. Catchphrases are found in the document, we used the catchphrases are gold standard for our summarization experiments. Citation sentences are found in later cases that cite the present case, we use citation sentences for summarization. Citation catchphrases are the catchphrases (where available) of both later cases that cite the present case, and older cases cited by the present case. Citation classes are indicated in the document, and indicate the type of treatment given to the cases cited by the present case.

**Attribute Information:**

Provide information about each attribute in your data set.

**Relevant Papers:**

[1] F. Galgani, P. Compton, and A. Hoffmann. Citation based summarisation of legal texts. In PRICAI 2012, volume LNCS 7458, pages 40â€“52. Springer, Heidelberg, 2012.   
[2] F. Galgani, P. Compton, and A. Hoffmann. Combining different summarization techniques for legal text. In Proceedings of the Workshop on Innovative Hybrid Approaches to the Processing of Textual Data, pages 115â€“123, Avignon, France, April 2012. Association for Computational Linguistics.   
[3] F. Galgani, P. Compton, and A. Hoffmann. Knowledge acquisition for categorization of legal case re- ports. In D. Richards and B. Kang, editors, PKAW 2012, volume LNAI 7457, pages 118â€“132. Springer, Heidelberg, 2012.   
[4] F. Galgani, P. Compton, and A. Hoffmann. Towards automatic generation of catchphrases for legal case reports. In the 13th International Conference on Intelligent Text Processing and Computational Linguistics, volume 7182 of Lecture Notes in Computer Science, pages 415â€“426, New Delhi, India, 2012. Springer Berlin Heidelberg.   
[5] F. Galgani and A. Hoffmann. Lexa: Towards automatic legal citation classification. In J. Li, editor, AI 2010: Advances in Artificial Intelligence, volume 6464 of Lecture Notes in Computer Science, pages 445 â€“454. Springer Berlin Heidelberg, 2010.