

Leximeter™... If it is text, it can be measured.

Operation and Benefits

Leximeter™ is a new product that integrates the measurement of document complexity, vocabulary analysis and writing style. This software package has a unique ability to compare documents or groups of documents to determine differences and similarities.

Leximeter™ analyzes the text of:

- ✧ emails
- ✧ forum posts
- ✧ blog entries
- ✧ business documents
- ✧ student papers
- ✧ ... and even novels

Leximeter™ uncovers similarities and differences in documents to help you make predictive and prescriptive analyses. Examples:

- ✧ winning proposals vs. losing proposals
- ✧ resumes as predictors of performance
- ✧ authorship of mysterious letters and emails

Leximeter™ is applicable to:

- ✧ Law and Law Enforcement
- ✧ Intelligence & Security
- ✧ Federal Agencies and Government Contracting
- ✧ Business correspondence with vendors and customers
- ✧ Academia and research

The results of the analyses are stored in Leximeter™'s database from which results can be extracted in both graphic and text formats.

Analyses and Comparisons

Under one roof, Leximeter™ combines the well known and many less well known text analysis algorithms, and scores the results in a way such that it is easy to see how and where documents or groups of documents differ and are alike. Leximeter™ is entirely dictionary driven, meaning that updates and changes to its knowledge of English do not require rebuilding the product.

Let's say your company has distributed a number of documents or proposals that are divided into two groups: Successful/accepted (*i.e.*, proposals that made it past the first round) and rejected. You suspect there is something about each document's content that influences which group it falls into.

The answer can be found by Leximeter™.

Under one roof using a single software package, Leximeter™ scores the results in a manner that makes it easy to distinguish differences and similarities.

The analyses include:

- ✧ Familiar estimates of document complexity, such as the [Gunning Fog Index](#) and the [Flesch-Kincaid](#) measures.
- ✧ Content word analysis, including a thorough inventory of the vocabulary used, with special treatment of Basic English, jargon, rare words, and contractions.
- ✧ [Function word](#) analysis, including an approach similar to the one that settled the disputed authorship of some essays in *The Federalist Papers*.¹
- ✧ Modern analysis of lexical bundles, both generic ("some of the", "on the other hand") and ones that arise in the documents you are analyzing.

¹Fung, Glenn, "The Disputed Federalist Papers: SVM feature selection via concave minimization," New York City, ACM Press, 2003.

The **Leximeter™** Family of Products

Researchers and any users with modest volumes of documents may want **Leximeter Lite™**, available in a software-as-a-service format with a convenient web interface.

For companies with sensitive documents or larger volumes, **Leximeter™** can be delivered pre-installed on a server, and all you will need to do is supply power and network.

Leximeter™ can also be installed directly in your environment, which allows you to make use of your corporate Oracle instance instead of using a separate database for **Leximeter™**.

Finally, you may wish to have us add specific functionality to **Leximeter™**, and **Leximeter™** is **designed to be extensible**.

For more information about purchasing a copy of **Leximeter™**, please call +1.724.GASLITE (+1.724.427.5483) US Eastern Time.

Technology

Leximeter™ is a computationally intensive program that takes full advantage of parallel processing. Multiple running processes support parsing several documents at once, and multiple threads within those processes support parallel execution of the analysis algorithms.

We used the most robust and stable technologies available today to construct **Leximeter™**.

- ✧ Network sockets, interprocess communication and mathematics libraries were selected from the **boost** project.

- ✧ The **flot** package was used for graphing the data.
- ✧ Database functionality in most cases is provided by **SQLite**, and **Oracle** integration is available as an option for existing Oracle customers.
- ✧ The web interface for **Leximeter Lite™** was written using **CodeIgniter**.
- ✧ All server side code was written in ISO and POSIX compliant C++, using the GNU compiler provided by **The Free Software Foundation**.
- ✧ **Leximeter™** runs on **Linux** or UNIX, and **Leximeter Lite™** uses the **Apache** web server.
- ✧ Energy efficient computers were purchased from **Apple**, and our refurbished Cisco routers, switches and firewalls from **DigitalHoop**.
- ✧ All documentation, including this paper, was created using the **L^AT_EX 2_ε** typesetting system.

Support and Services

Leximeter™ was entirely written by Digital Gaslight, Inc. in the United States of America, and the software is not subject to export restrictions.

All licenses for **Leximeter™** include full product support from Digital Gaslight, Inc., all upgrades and patches as they become available, and access to the customer forum where your questions will be answered in writing.

For large customers, on site training and consultation is available.