J23   Detecting Backdated Documents Through Line Layout Approaches to Font Identification

Thomas W. Phinney, MS, MBA, Portland, OR 97206*

Learning Overview: After attending this presentation, attendees will learn the utility of font identification in detecting backdated documents. Attendees will learn about font dating and its use in detecting backdated documents, as well as both traditional methods of font identification and newer methods based on glyph spacing, which are suited to automation and capable of yielding results even with very poor document reproduction.

Impact on the Forensic Science Community: This presentation will impact the forensic science community by not only covering widely used tools but also giving attendees new tools for detecting document backdating and identifying an area for more research.

How Font Identification Can Demonstrate Forgery: Identifying a font used in a document, when combined with information about when the font was designed and made available to the public, can be supportive or conclusive in proving that the document was not created on a date claimed. This can not only prove backdating, but in cases where the purported author was deceased before the font was available, can be evidence of forgery.

Traditional Methods and Resources for Font Identification: A variety of books, manuals, and web sites provide visual references and information on font classification and earmarks. Many of these approaches require some degree of expertise for reliable results.

Complications for Font Identification and Timelines: Although a font may be officially released on a given date, that does not mean it was completely unavailable before that date. Especially for fonts bundled with operating systems and office suites, the availability of pre-release and beta versions can complicate the story by creating a gray area of time, when a new font was not yet used by the average person, but still available to some people outside the company. Calibri illustrates this issue, with its increasing availability to outside testers for a period of almost two years before its formal and full release with Windows® Vista and Office 2007.

Another complication, for at least some common typefaces, is the existence of (later) lookalike fonts, sometimes called "clones" or "knockoffs" of the original. This generates the potential for false negatives, or less commonly, false positives.

Line Layout Methods for Font Identification: Except for traditional typewriters and fonts of that same style (such as Courier, Prestige Elite, and Consolas), wherein all letters take the same amount of space, most fonts are "proportional": each letter is allocated a different amount of space. Aside from some proportionally spacing typewriters and a few fonts created to imitate the spacing of earlier fonts, the spacing of fonts is generally unique, like a fingerprint.

This spacing is revealed by cumulative character width, especially with non-justified (ragged right) text. Even with a very poor document reproduction, one can use this method by comparing line breaks and especially ragged line endings. Any large-enough sample of such lines is like a fingerprint: the relationship of where each ragged line ends, relative to the adjacent lines, uniquely identifies a particular proportional font. Because this method is based on *relative* line endings, it does not require duplicating the original document and is unaffected by distortions that can occur when documents go through copying, faxing, scanning, etc.—even multiple generations of reproduction.

An employee of Adobe® at the time made this observation in 2004, in regard to the Bush National Guard memos. This concept, when applied in an automated way by a computer program, became the basis of a patent application. The concept was subject to internal peer review, and the patent was granted by the United States Patent & Trademark office and expires in September 9, 2025).

That this technique can differentiate different fonts has been thoroughly demonstrated on a small scale, with handfuls of fonts at a time. However, no testing has been done on a larger scale, to further validate the methodology or determine the statistical reliability of the method. Such testing can be done manually; if automated, it would presumably require either licensing the patent or waiting for patent expiry in 2025.

The presenter has no financial interest in Adobe® Systems nor in licensing the patent.

Forgery Detection, Font Identification, Document Authentication

*Presenting Author