

THE AMAZING SCOREIT! TECHNOLOGY: The Story Behind the Science

How Inkubate’s Algorithms Work to Identify Your Writing Style

Inkubate’s technology is within the exciting and advanced field of “natural language processing.” When analyzing writing style, we are referring to the elemental building blocks and stylistic constructs of written communications. This is different and more meaningful than the concept of “writing style” that you may have studied in the classroom. For our purposes, your “writing style” is determined by analyzing and evaluating thousands of distinct patterns of writing choices that you have developed over the course of your lifetime. Your distinctive style of writing is a function of a number of factors including your gender, education, nationality, cultural influences, and personality traits. By analyzing your manuscript, or even a chapter or two of it, Inkubate’s ScoreIt!™ can very accurately map your writing style “fingerprint.” By applying this same science to all of the many authors in our corpus and, by comparing your writing style fingerprint to theirs, Inkubate’s ScoreIt!™ tool reports back to you the writers to whom you are best matched. ScoreIt!™ answers the question: “Who do I Write Like?”

ScoreIt!™ aggregates thousands of proprietary writing style patterns into four major “families” to help you to more easily understand your personal writing style. These four families are as follows:

Major Writing Style Feature Families	
Vocabulary	The vocabulary and range used by the author to express ideas and the unique word combinations that describe the author’s literary “voice”
Expressive Complexity	The choice and distribution of word lengths. A large and complex vocabulary, for example, includes a variety of short and long words. Collectively, these word choices reflect an author’s distinctive expressive style.
Grammar	Describing complex human relationships typically requires more intricate grammar than describing a simple, action-based narrative. The range of word choices related to these “parts-of-speech” and the number and arrangement of prepositional phrases used reflect an author’s grammatical style.
Tonal Quality	This feature analyzes, among other aspects, the author’s use of “function words” such as “the”, “a/an” and “of”. Their use describes the functions of other words in a sentence and their frequency is a key characteristic of writing style.

It is the overall mapping of these “features” and their intersections with those of the writers in our corpus, that allows us to evaluate and identify similar writing styles. Our corpus contains representative examples of virtually all the generally recognized fiction genre. By example, but by no means exhaustive, some of the major genre on this list include

romance, fantasy, mystery, western, historical, and sci-fi. While Inkubate's corpus is constantly evolving, most of the authors who published a bestselling book during the last 10 years, as well as many of the classic authors from the late 19th century and early part of the 20th century, are represented.

Because ScoreIt!™ answers the intriguing question of “who do I write like?”, it is common that many of the authors who you write like will be writing within a genre close to yours. However, it is also common that some of the matched authors will be writing in other genres less closely aligned. That provides fascinating feedback and important insight that can help you identify who your audience may be and how you might target a wider audience through careful metadata selection. Genre classification and marketing keywords are examples of critically important metadata that enable your target audience to discover your book.

Here, the analogy to the music service technologies previously mentioned is instructive. For instance, a person who loves the musical group *Queen* would likely be attracted to other musicians whom, at a macro level, could be easily classified into a variety of music genres including: rock, heavy metal, rock opera and perhaps others. But that person might also like classical music, jazz, opera or show tunes. Why is that?

It is because music, like writing, is a form of communication. Your preferences are influenced by the sum total of your exposure to the vast variety of available communication content (in this case, music) and your unique and fluid reactions to that content. As you learn what you like, you learn to identify other new content that pleases you. While genre classifications are useful and help us to generally describe our likes and dislikes to each other, each piece of music is unique – just like each book. Your brain does not think about genre, it simply likes what it likes. It is limiting to think about your likes and dislikes by looking at genre classification alone.

In the case of writing style, Inkubate leverages our ability to analyze why two unique writers are similar without considering the direct influence of genre, plot or “story arc.” This provides a new and valuable metric – writing style – that can be used to help writers find readers and readers to find writers. Furthermore, identifying writer similarities based upon these techniques has been proven by us to have important relevance and value in the marketplace. Consider the following well-known story:

HOW SCOREIT! TECHNOLOGY SOLVED THE CASE OF THE MYSTERY AUTHOR

A Case Study: *The Cuckoo's Calling* by Robert Galbraith

In 2013, a book was published by the British publishing industry called “The Cuckoo’s Calling” by Robert Galbraith. The book was unusually accomplished for an apparent first-time author and, surprisingly, the book began to rise quickly on the bestsellers lists in a very short period of time. As the book quickly became popular, people started asking questions.

Though not impossible, it was unusual that no one in the publishing industry had ever heard of this author. Additionally, no history of such a person could be found in public records and the author’s biography was full of “facts” that could not be confirmed as well as numerous inconsistencies. Nobody could identify any previous manuscript submissions to any agents and there was no evidence that Robert Galbraith had ever self-published any titles. There was even a complete lack of any social media presence prior to the publication of his first book. Suspicion quickly arose that Robert Galbraith was, in fact, a “pen name” for somebody else.

London’s *Sunday Times* took the lead in this investigation, determined to discover the true identity of Robert Galbraith. “Who really is Robert Galbraith?” – The *Sunday Times* asked Inkubate’s data scientists for help in solving this mystery.

As suspicions grew, a tip from a reliable source received by a *Sunday Times* reporter claimed that J.K. Rowling had secretly written a novel under a pen name. Other corroborating evidence included the fact that Galbraith’s agent and editor were none other than J.K. Rowling’s. But when confronted with these facts, Rowling and her representatives repeatedly stonewalled, denied any such connection and furthermore, they threatened a lawsuit if the *Sunday Times* continued to pursue its “unfounded” allegations.

Inkubate’s data scientists brought the same technologies upon which ScoreIt!™ is based to bear on this question. By fingerprinting the writing style of Robert Galbraith, based upon his novel *The Cuckoo’s Calling*, and comparing that to J. K. Rowling’s writing style fingerprint derived from one of her acknowledged published works, the result was a near perfect match. To quote our science team (with a little tongue in cheek!):

“Though we cannot tell you the name of the author who wrote *The Cuckoos Calling*, that person writes identically to J. K. Rowling!”

Armed with this, the *Sunday Times* approached Rowling’s representatives yet again. Faced with the overwhelming evidence collected, but most importantly, the empirical analysis of writing style, they finally had to admit that *The Cuckoo’s Calling* was in fact Rowling’s work. As suspected, she had hoped that by publishing under a pseudonym, she would get market feedback and objective reviews about her new work without the influence of her prior reputation and any biased expectations. As you may know, *The*

Cuckoo's Calling was the first in a series of books about a former military man written in the crime fiction space – a genre very different indeed from that of J.K. Rowling's Harry Potter fantasy series.

Two important conclusions are evident in this case study. First, writing style can be analyzed and “fingerprinted” and then compared to find the best matches with other authors and their “fingerprints.” Second, writing style matters to readers, as evidenced by the rapid rise in popularity of a book by an author that hid their true identity in order to explore new genre. Though her subject matter was very different, her underlying writing style, the part of her presence that she could not easily disguise, came through to her readers.

The “outing” of J. K Rowling is captured in an article in [Scientific American](#), by Dr. Patrick Juola who is Inkubate's chief data scientist and forensic linguist. If you are interested, the entire article, it can be found [here](#).