

eliminated. After these manual process steps are completed, a SFP count is created by counting verbs and assigning the SiSE weight to each verb instance.

Recent experience at the Department of Homeland Security (DHS)³ as well as the National Geospatial-Intelligence Agency (NGA) shows that a SFP count can successfully be created using SiSE from a single document, such as a Concept of Operations or Lean Business Case.

Natural Language Processing

While all the sizing methods discussed have their own merits, they require some level of expertise in software sizing, as well as the time and manpower to create the metric. In situations where only a single sizing number is needed (e.g., for a single program), the schedule and effort requirement is likely not an issue. But for organizations that need a software sizing estimate on a portfolio of multiple programs, an automated solution offers some benefit. Additionally, an automated solution to software sizing is based on a single algorithm that can be applied consistently to many estimates. As such, consistency can be achieved without the need for published standards or a certification program. For these reasons, a solution based on Natural Language Processing (NLP) offers some significant advantages.

NLP Overview

Though rarely applied in cost estimating, NLP is not a new concept. It is used today in many different applications, such as: classifying emails as spam, autocorrect, spellcheck, and foreign language translation. For data analysts, NLP offers a compelling capability: converting unstructured text into structured data. When narrative text can be expressed as numbers, it opens the door to a wide variety of statistical and data-driven approaches that would otherwise be unavailable.

³ Let's Go Agile: Data-Driven Agile Software Cost and Schedule Models Derived from DHS Projects, <https://www.iceaaonline.com/wp-content/uploads/2022/06/SA09-Rosa-Lets-Go-Agile.pdf>

