

## Sample of Level 2 English Editing

Field of research: Imaging and Informatics

Using UMLS to Construct a Generalized Hierarchical Concept-based

Dictionary of Brain Function for Information Extraction from fMRI Literature

### 1. Introduction

Functional magnetic resonance imaging (fMRI), which measures [the](#) blood metabolism of neural activities [1], is a non-invasive approach for studying human brain function. Over the past 15 years, fMRI has become extremely popular among neuroscientists. As a result, [a](#) profusion ~~e-number~~ of fMRI studies are published every year, in numerous journals. Hence, it is becoming difficult for researchers to keep up with the literature by reading papers, as this activity is extremely time-consuming and labor-intensive. Therefore, [automated](#) ~~ie-analysis and understanding of~~ unstructured texts using a text mining technique [such as that](#); ~~as~~ proposed here, ~~,~~ may be of great help to scientists.

The main challenge in text mining, especially in the field of neuroscience, is the [problem of](#) data entry ~~problem~~. Currently, [the](#) automatic retrieval ~~ing~~ and extraction ~~of~~ ~~ng~~ large amounts of data for analysis is still difficult, and most databases rely on [inputting](#) information ~~manuall~~ ~~input manually~~. For instance, the first annotated database for published neuroimaging studies was BrainMap [2, 3], [the](#) ~~its~~ interface [of which](#) allowed researchers to search by querying

**Comment [SM1]:** CHECK: Reference to “understanding” was unclear; in fact this is a human function rather than a computerized function, so sits oddly with the automated analysis; thus have left this out. Reference to helping scientists gets the point across anyway.

experimental parameters, bibliographical details, or specialized locations in Talairach space [4]. However, the scale of BrainMap was limited; currently, only 61 behavior related terms or function names, and 72 experimental paradigms, exist-isare—available in the database. Such a limited scale may be inevitable, because BrainMap relies on researchers to input data manually. Thee lack of automatic processing severely limits the scope of the database and reduces its usefulness.

The goal of the work presented in thise paper was to build an automatic information extraction (IE) tool that can extract terms related to brain anatomy, function, and experimental tasks from the fMRI literature. Such an approach should allows for the processing of a large amount of text data in a relatively short period and couldcan overcome the shortcomingsfalls of the manual entry approach, ases discussed above. To demonstrate the possible applications of this system, we constructed the-a brain-function co-occurrence association model to assist with identifying terms associated with study human brain function.

IE is a method that allows automatic recognition of meaningful words or phrases from free text. A variety of IE methods have been applied to bioinformatics, either by-in the dictionary based [5] or the rule based approaches [6, 7], in applications including detection of disease, protein and gene names [8-11]. IE methods have also been used for identifying relationships between different terms —, for

**Comment [SM2]:** CHECK: is this just an abbreviation or is IE a recognised method and approach? If the latter, perhaps you should extend into giving a more technical account here, for the sake of clarity (see reviewer comments in general).

**Comment [SM3]:** IDEA: In this version of the paper at least, you only state that manual input resulted in a reduced scale in Brain Map. Presumably you are implicitly referring to other issues in this passage, but not you don't cite them. I suggest you do so: eg time consuming, subjective judgment, involves targeted training, slow, etc. etc. Also I have made "manual entry approaches" in the singular because you only seem to mention one approach in general?

**Comment [SM4]:** CHECK: What does this mean? What is "free text"? Reviewers want unfamiliar terms explained; will they know this term? Likewise, in the first par of your intro you refer to "unstructured text". I suggest you use just one term (if these 2 terms mean the same thing?) and then clearly define what you mean.

example, protein-protein interactions [12-15]. While much progress has been made in applying IE to bioinformatics, unfortunately, it is difficult to ~~directly~~ apply these achievements directly to a different domain, such as neuroscience, as the textual features subject to IE are highly domain-specific.