

# Parsing Free Text Nursing Notes

William J. Long

MIT Lab for Computer Science, Cambridge, MA, USA

*Parsing nursing notes requires tokenization, recognition of special forms, abbreviation expansion, and classification in the context of identified sections.*

## Introduction

Nursing notes entered as free text are often the best way to track the progress of a patient from online data. However, they present a challenge because they are highly telegraphic, with many abbreviations both common and invented on the spot and with punctuation serving as abbreviations. Furthermore, many of the abbreviations and forms are dependent on context for their interpretation. The result is that it is necessary to provide a level of preparsing before the usual word, phrase, and sentence analysis can take place.

We encountered this problem in the context of tracking the progress of patients in the ICU where nursing notes are added to the patient record at the end of each shift and occasionally when important events take place, but the problems and approach are more general.

## Approach

Several of the challenges are illustrated by the following line from a note:  
NAUSEA-NOVOMITING.ACTION-ZOFRAN  
2MG IV GIVEN X1.

The tasks can be broken down as follows:

1. **Tokenization:** Determining how to divide the text into appropriate units. The periods mark the ends of phrases even though spaces are missing, except when it is part of a floating point number. Usually, punctuation marks can be used to determine tokens. An additional problem is missing spaces as in the NOVOMITING. This requires recognition of words and is handled as a spelling correction problem.

2. **Recognition of special forms:** Forms such as X1 (repetitions), 2mg (value in units), 8/20 (date), 120/70 (pressure), 102-104 (range), 90's (range), w/ (abbreviation for with), dc'd (past tense) require flexible recognition that

allows a class of entities to match the form. Some of these have additional constraints or context, such as the difference between a date and a pressure. Punctuation can serve as an abbreviation, such as "+" for present or "^" for increased.

3. **Expansion of abbreviations:** Many of the abbreviations encountered are commonly used, such as "noc" for "at night" but many are dependent on context as the "MG" above is milligrams but without the "2", it might be magnesium. Some abbreviations are simply the first few letters of a word common in the context of the patient, such as "chol" for cholesterol. Others represent a phrase, such as "Tmax" for maximum temperature.

4. **Spelling correction:** Since the notes are entered rapidly by hand there are plenty of misspellings. Correcting these requires a dictionary with the medical vocabulary that arises in the clinical context.

5. **Section recognition:** Most nursing notes have markers to indicate sections, either of the SOAP form (as the ACTION above) or as a review of systems. Determining the sections provides important context for resolving ambiguous abbreviations. For example, "bs" is breath sounds in the respiration section, bowel sounds in the GI section, and blood sugar in the context of lab results or a diabetic.

Once these tasks have been completed, the text is ready for the application of a medical parser.

## Conclusion

This approach makes it possible to process the nursing notes and use the information in them to track the changes in symptoms, medications, procedures, and significant events – the information necessary to interpret the clinical monitoring data.

## Acknowledgements

This research has been supported by contract 290-00-0020 from the Agency for Healthcare Research and Quality. The author thanks Dr. Roger Mark for access to the data.