ABSTRACT

In this chapter, we describe the development and application of language technology for intelligent information access to the content of digitized cultural heritage collections in the form of Swedish classical literary works. This technology offers sophisticated and flexible support functions to literary scholars and researchers. We focus on one kind of text processing technology (named entity recognition) and one research field (literary onomastics), but we try to argue that the techniques involved are quite general and can be further developed in a number of directions. This way, we aim at supporting the users of digitized literature collections with tools that enable semantic search, browsing and indexing of texts. In this sense, we offer new ways for exploring the large volumes of literary texts being made available through national cultural heritage digitization projects.

DOI: 10.4018/978-1-60566-932-8.ch003
INTRODUCTION

Literature can be studied in a number of different ways and from various perspectives, but text analysis – in a wide sense – will probably always make up a central component of literature studies. In this day and age, the computer has become an indispensable tool in many kinds of text analysis, such as in linguistics and in information access, to mention two fields. However, the potential of the computer as a text analysis tool in literature studies is arguably largely untapped (Bradley 2005; Juola 2008).

Literature is now increasingly available in electronic form. Modern literature is born digital as a matter of course (but may not always be available in this form to research), while older literature is being computerized apace as part of national cultural heritage preservation efforts. Thus, there is no shortage of literary works in digital form. We can see this as an opportunity to explore how text analytical tools that have turned out to be useful in other fields could be put to effective use also in literature studies. In our case, we are interested in developing and exploring computer tools based on language technology, which is our primary field of expertise. Here, a reasonable approach would be to identify some particular component or subfield of text analysis in literature studies having requirements which would match the capabilities of some mature language technology, and further see how this could be packaged up in a way that would be of help to literary scholars.

In this chapter, we argue that the field of literary onomastics is part of literature studies. We also point out that there is a mature language technology which can be ancillary to text analysis in this field, namely the technology of named entity recognition.

In the next section, we briefly describe literary onomastics and named entity recognition. In Section 3, we give an account of our recent work on automatically providing a large number of digitized classical Swedish literary works with named entity annotations and making these annotations accessible to users through a search and browsing interface. In this account, we try to cater both to those readers who are mostly interested in the implications of our work for literary and other Humanities scholarship, and to those who are curious about the technology involved. Section 4 outlines some future directions in which we would like to continue our research, specifically intelligent information access as applied to literary and historical texts. Finally, Section 5 offers some conclusions to this chapter.

DEFINING THE AREA

Literary onomastics is a field of inquiry where literature is seen through the names appearing in literary texts. Specific topics may comprise studies of the etymology or
Addressing the “Commitment Problem”: Driving Long-Term Persistent Focus in Self-Discovery Learning (A Thought Experiment)
www.igi-global.com/chapter/addressing-commitment-problem/61308?camid=4v1a