Named Entity Extraction

What does it do?   Given a text, it gives you a list of people and things mentioned in it. The most popular implementation extracts names of people, places, and organizations.

What's so great about it?   Sometimes you don't really want to work with text: you just want to hoover out specific details. Maybe you want to make a map of places mentioned in a book. Maybe you want to make a social network based on who shows up in social registers.

What choices do I have to make?   You'll almost certainly be using Stanford's software, in which case you need to choose:

1. Which *model* to use. Different languages have different ways of training places.
2. Whether to build a new model on your own. This is a huge investment, but can be useful if your texts are quite different from the contemporary news stories the English-language Stanford Texts for trained on.

What software should I use? Are there alternate implementations?   Use the Stanford Natural Language Processing Group's Named Entity Recognizer. It's likely that neural network parsing will displace Stanford's rules at some point, but this remains far and away the most widely used implementation.