

Appendix 5

Named Entities: Current Definitions

This appendix provides a non-exhaustive list of different definitions applied to named entities. These descriptions are taken from research program reports, general works on NLP, publications relating specifically to named entities, online encyclopedias and the Websites of companies which offer information processing services. This list provides a good overview of the general discourse concerning named entities, although readers should remember that definitions may refer to relatively different contexts and applications.

In research programs:

MUC-7: “On the level of entity extraction, Named Entities (NE) were defined as proper names and quantities of interest. Person, organization, and location names were marked as well as dates, times, percentages, and monetary amounts”[CHI 98].

“The Named Entity task consists of three subtasks (entity names, temporal expressions, number expressions). The expressions to be annotated are ‘unique identifiers’ of entities (organization, persons, locations), times (dates, times), and quantities (monetary values, percentages)” [CHI 97].

CoNNL: “Named entities are phrases that contain the names of persons, organizations and locations” [TJO 03].

ESTER: “While no standard definition exists, NE may be said to constitute a particular type of lexical unit referring to a real-world entity in certain

specific domains, notably the human, social, political, economic and geographic domains, and which have a name (typically a proper noun or an acronym)” [MEU 04].

Quaero: “Named entities traditionally include three broad classes: names, quantities, and dates and durations. We will consider named entities in the context of information retrieval (entities, relationships) where they are used to constitute a knowledge base” [ROS 11].

In general studies and works on named entity recognition:

B. DAILLE AND E. MORIN: “The notion of named entity represents a much broader category than that of proper nouns (...), as it includes temporal or numeric expressions, diseases or drugs” [DAI 00].

S. SEKINE: “The names of particular things or classes, and numeric expressions is regarded as an important component technology for many NLP applications (...) In this paper, the term Named Entity includes names (which is the narrow sense of Named Entity) and numeric expressions. The definition of this Named Entity is not simple, but, intuitively, this is a class that people are often willing to know in newspaper articles” [SEK 02].

N. FRIBURGER: “It is, in fact, difficult to differentiate between proper nouns and other nouns; there is a degree of continuity between the set of proper nouns and the set of common nouns. Computer scientists working in the field of information retrieval have taken a pragmatic approach to the issue, defining the notion of *named entities* to cover all elements of language which may be defined by a reference: proper nouns in both the traditional and broader senses, but also expressions of time and quantities” [FRI 02].

D. WEISSENBACHER: “To give a provisional definition, a named entity is a syntagm which refers to a single object in a presumed reality. Proper nouns, company names, locations, dates, etc. fall under this definition. (...) In this work, we will consider that an NE is an unambiguous name for an essential participant in the event being modeled. Note that the first definition is included in the second; to denote a single individual, we must simply establish the concept of being this individual” [WEI 03].

- T. POIBEAU:** “The term “named entity” is traditionally used to refer to the set of person, organization and location names encountered in a given text. Further syntagms, such as dates, monetary units or percentages, are often included and detected using the same techniques, based on local grammars. In this chapter, we will only consider named entities in the strictest sense of the term (...)” [POI 03].
- M.R. VICENTE:** “Named entity *is the notion used in NLP to denote monoreferential discursive elements, partly coinciding with the set of proper nouns (note that the notion of named entities concerns proper nouns, but also dates and measurements), and following established syntactic patterns*” [VIC 05].
- M. TRAN:** “In MUC-6, proper nouns, dates and figures were grouped together under the term “named entities” [TRA 06a].
- GALIBERT, ROSSET:** “Mono or multi-word expression belonging to a potentially interesting class for an application” [GAL 11].
- D. NOUVEL:** “Named entities, when disambiguated, appear to denote mental objects in a *stable* manner, allowing the creation of a logical representation” [NOU 12a].
- A. BETREMIEUX:** “Named Entity is a generic term used to categorize a certain number of textual objects encountered in a document”.
- D. NADEAU AND S. SEKINE:** “The word ‘Named’ aims to restrict [Named Entities] to only those entities for which one or many rigid designators, as defined by S. Kripke, stands for the referent. [...] Rigid designators include proper names as well as certain natural kind terms like biological species and substances. There is a general agreement in the NERC community about the inclusion of temporal expressions and some numerical expressions such as amounts of money and other types of units. While some instances of these types are good examples of rigid designators (e.g. the year 2001 is the 2001st year of the Gregorian calendar) there are also many invalid ones (e.g. in June refers to the month of an undefined year past June, this June, June 2020, etc.). It is arguable that the NE definition is loosened in such cases for practical reasons” [SEK 07].

D. JURAWSKY AND J.H MARTIN: “A named entity is everything ’that can be referred to with a proper name” [JUR 08], p.761.

“The notion of named entity is commonly extended to include things that are not entities per se, but nevertheless have practical importance and do have characteristic signatures that signal their presence” [JUR 08], p. 762.

From encyclopedias and company Websites:

WIKIPEDIA, article *Named Entity Recognition*: “Named-entity recognition [...] is a subtask of information extraction that seeks to locate and classify elements in text into pre-defined categories such as the names of persons, organizations, locations, expressions of times, quantities, monetary values, percentages, etc.”.

BUSINESS ANALYTICS: “In data mining, a named entity is a phrase that clearly identifies one item from a set of other items that have similar attributes. Examples of named entities are first and last names, geographic locations, ages, addresses, phone numbers, companies and addresses. Named entities are often mined for marketing initiatives¹”.

ALCHEMYAPI: “Named entities specify things such as persons, places and organizations. AlchemyAPI’s named entity extraction is capable of identifying people, companies, organizations, cities, geographic features and other typed entities from your HTML, text or web-based content²”.

1 <http://searchbusinessanalytics.techtarget.com/definition/named-entity1>.

2 <http://www.alchemyapi.com/products/alchemylanguage/entity-extraction/>.