

Tato bakalářská práce se zabývá rozpoznáváním pojmenovaných entit a jeho aplikací pro detekci phishingu v emailové korespondenci. Rozpoznávání pojmenovaných entit je klasifikační úloha, jejíž cílem je extrahovat informace z textu klasifikovaného do předem definované množiny kategorií (pojmenované entity) jako například názvy organizací, míst či osob. Dále práce popisuje různé přístupy k rozpoznávání pojmenovaných entit, od jednoduchého využití neuronových sítí k momentálně nejmodernějším architekturám. Nejrozšířenější knihovny a jejich modely týkající se rozpoznávání pojmenovaných entit jsou navzájem porovnávány z pohledu výpočetního výkonu a přesnosti predikcí na veřejně dostupném Enron email datasetu. Rozdíly v nalezených pojmenovaných entitách mezi pozitivními (včetně phishingových) a negativními emaily jsou určeny na proprietárním datasetu. Na závěr je tento dataset použit pro experiment, kde proces klasifikace phishingových e-mailů je obohacen o pojmenované entity k cílem dojít k závěru, zda jsou pojmenované entity užitečné pro klasifikátor ke zlepšení prediktivního výkonu. Na základě výsledků experimentů byla pozorována znatelná odlišnost v pozitivních a negativních emailech. Nicméně v experimentu týkající se klasifikace phishingových emailů nad tímto datasetem se došlo k závěru, že pojmenované entity nepřináší přesvědčivý benefit. Je však možné, že s jiným datasetem by se mohly odvozené závěry lišit.