

Lingvistický diskurz se zabývá významem delších kusů textu, od vět po celé dokumenty, mohl by se však uplatnit i v úlohách získávání informací z textu, např. vyhledávání dokumentů či jejich sumarizace. Cílem této práce je uplatnění informací o stavbě diskurzu psaného textu pro potřeby získávání znalostí. Jedná se o první pokus, který se snaží skloubit tyto dva velice odlišné obory, a jeho ambicí je tak připravit základ pro tento způsob získávání znalostí. Náš postup spočívá v použití metod neřízeného strojového učení k analýze diskurzních vztahů a jejich následném modelování pomocí vzorových struktur z formální konceptuální analýzy. Naši metodu jsme aplikovali na korpus lékařských článků z databáze PubMed. Tyto lékařské texty potom obohacujeme o koncepty z metathesauru UMLS, které jsou kombinovány s daty ze sémantické sítě UMLS, která fungují jako ontologie ve vzorových strukturách. Naše výsledky ukazují, že i přes vysokou úroveň šumu je naše metoda slibná a bylo by možné ji aplikovat i na jiné domény.