

Jazyk XML je stále důležitější formát pro uchování a výměnu dat. Pro korektní výměnu dat, jejich efektivní ukládání a zpracování je nutné znát jejich strukturu. Stále však existuje a je vytvářeno mnoho dokumentů bez popisu jejich struktury. Tato práce se zabývá možností automatického generování schémat popisujících strukturu na základě dané množiny dokumentů. Presentovaný algoritmus rozděluje elementy dokumentu do skupin podle podobnosti. Pro každou skupinu je vygenerován regulární výraz co nejlépe popisující strukturu vstupních elementů s využitím možností jazyka XML Schema.