

## Errata k bakalářské práci

### Metody extrakce víceslovných výrazů z textu

Už po odevzdání práce k tisku bylo zjištěno, že práce obsahuje následující chyby (resp. neúplné informace):

#### 1. Zjišťování kontextu (kapitola 4.6)

V práci je uvedeno, že se pro každý N-gram rozlišují dva typy kontextu – široký a úzký. Dále je uvedeno, že kontext se zjišťuje pro všechny „hvězdičkové N-gramy“ (s výjimkou stupně nula). To ovšem platí pouze v případě širokého kontextu – úzký kontext se zjišťuje pouze pro **úplné** N-gramy.

#### 2. Uložení kontextu (kapitola 6.3.5), Přetečení čítačů (kapitola 6.3.6)

V práci chybí informace o tom, že pro rozlišení zda dané slovo v kontextu (ve smyslu jeho uložení v paměti – v rámci kontejneru typu *vector*) je součástí úzkého nebo širokého kontextu, je použit nejvýznamnější bit čítače výskytu. Maximální hodnota čítače výskytu pro slova v kontextu je proto  $2^{31}-1$  (a ne  $2^{32}-1$ , jak je uvedeno v práci).