

Fórum:Testy2/Příklady realizace

Automatická tvorba testů na 3. LF UK

V rámci integrovaného předmětu *Buněčné základy medicíny* vytvořil v roce 2011 tým učitelů 3. LF UK systém pro automatizovanou tvorbu testů z předem připravených souborů tvrzení.

Použité testy mají formální strukturu MTF testů, tedy jde o soubor tvrzení, u nichž studenti určují, zda jsou pravdivá či nikoli. Z hlediska teorie testování se nejedná o ideální formu, avšak tato volba byla výsledkem kompromisu mezi vhodností formátu a složitostí přípravy testů. Omezení MTF formátu jsou částečně kompenzována druhou částí testu, která se skládá z úloh typu *short answer*. Konkrétní MTF testy jsou náhodně losovány z textových souborů tvrzení, z nichž každé je označeno číslem oboru, kurzu (podjednotka integrovaného předmětu), přednášky a písmenem označujícím obtížnost daného tvrzení.

Jednotlivé úlohy v MTF testu jsou složeny vždy ze čtyř tvrzení, které spojuje tématická příbuznost; v případě tohoto předmětu jde vždy o čtyři tvrzení k tématu jedné přednášky. Kmen každé úlohy je naprosto generický, např. *Která z následujících tvrzení jsou správná?* Aby byla zaručena srovnatelná obtížnost testů v různých termínech, jsou jednotlivá tvrzení označena relativní obtížností (A, B nebo C), kterou určí přednášející daného tématu po případné dohodě s dalšími učiteli daného tématu či oboru. Jedna z možností nastavení škály obtížnosti je následující. Otázku C musí znát každý student, jde o tzv. bazál. Obtížnost A je pro výborné studenty, kteří aspirují na nejlepší známku, často se jedná o tvrzení týkající se detailů či vyžadující studium z dodatečných materiálů. Tvrzení s obtížností B jsou potom někde mezi těmito dvěma extrémy – znalost alespoň poloviny z nich (spolu s tvrzeními C) je minimálním předpokladem úspěšného složení zkoušky. V popisovaném případě obsahuje každá úloha jedno tvrzení C, dvě B a jedno A. Tyto úrovně obtížnosti mohou být zhruba namapovány na Angoffovy pravděpodobnosti takto: C=1,0; B=0,75 a $0,5 < A < 0,75$. Označení obtížnosti se používá pouze pro vytvoření testu, studenti je nevidí.

Takovéto rozdělení obtížností za předpokladu jejich realistického přiřazení navíc umožňuje nastavit *a priori* minimální bodový zisk nutný pro složení zkoušky. Student, který zná všechna tvrzení obtížnosti C a polovinu tvrzení B tak získá 50 % bodů a vzhledem k pravděpodobnosti úspěšného natipování má slušnou pravděpodobnost dosáhnout 75 %, což je v daném případě hodnota minimálního počtu bodů pro úspěšné zvládnutí zkoušky. Vzhledem k celkovému počtu 396 tvrzení v tomto testu je pravděpodobnost získání 75 % při znalosti 50 % tvrzení asi 0.53 a prudce stoupá s dodatečnými tvrzeními, které daný student zná.

Jelikož tvrzení C jsou učiteli-experty považována za esenciální, při vyhodnocování testu je jejich neznalost navíc výrazně trestána anulováním celého bodového zisku z dané otázky (tedy souboru čtyř tvrzení). Tak se výrazně snižuje pravděpodobnost, že uspěje student, který sice zná tzv. špeky, ale nemá patřičné základní pochopení či vědomosti.

Testy jsou generovány pomocí skriptu v Perlu podle předem daných pravidel určujících počet úloh z daných oborů a témat. Vygenerované testy jsou pak importovány do programu, který vytvoří varianty a po napsání testu umožní naskenovat odpovědní formuláře a test automaticky opravit.