**Маргарита Брус, Олександр Костіков, Сергій Загребельний**

**(Краматорськ, Україна)**

**ОЦІНКА ЯКОСТІ ТЕСТОВИХ ЗАВДАНЬВ MOODLE**

У зв'язку з впровадженням дистанційного навчання на кафедрі інформатики і інженерної графіки Донбаської державної машинобудівної академії виникає проблема оцінки якості тестових завдань. Тому перш за все треба оцінити валідність розроблених тестів. Найбільш вдале визначення валідності педагогічного тесту дав Анастазі [3]. Валідність тесту – це “поняття, яке визначає, що вимірює тест і наскільки якісно це здійснюється”[4]. Невалідність тесту означає, що одержані результати не вимірюють того, що було визначено метою тестування. У нашому випадку це означає, що оцінки, які отримав студент в результаті комп'ютерного тестуванняне відповідають його справжньому рівню знань .Аналіз валідності результатів базується на спеціально розроблених методиках зіставлення з критеріями аналізу об’єктивності результатів, в основі яких – оцінка міри відхилення від нормального закону результатів тестування та процедура їх нормалізації [5]. Ступінь відхилення від нормального закону визначається за такими параметрами, як асиметрія Аs і ексцес Еx , та встановленням їх статистичної достовірності за t-критерієм Стьюдента. Залежно від значення та знаку асиметрії або ексцесу результати можна поділити на дев’ять загальних груп, що дають змогу визначити критерії прийняття рішення щодо валідності процесу вимірювання та наявності або відсутності статистично достовірної дії факторів, що впливають на результати вимірювання.

Залежно від значення і знаків асиметрії та ексцесу закону розподілу можна виділити дев’ять таких критеріїв[3]:

1) Тест є валідним, якщо асиметрія та ексцес закону розподілу дорівнюють нулю ( As = 0, Ex = 0 ).

2) Якщо асиметрія є від’ємною, а ексцес додатним (As< 0, Ex> 0), то тест є невалідним (легкий і з малою розподільною здатністю). Це результат дії кількох факторів: – тест сконструйовано з невалідних тестових завдань (він мусить валідизуватися шляхом їх ускладнення та збільшення їх розподільної здатності); – процедура тестування є невалідною.

3) Якщо асиметрія та ексцес є від’ємними ( As< 0, Ex< 0), то тест є легким, але з великою розподільною здатністю. Його валідизація полягає в ускладненні тестових завдань.

4) Якщо асиметрія є від’ємною, а ексцес відсутній (As< 0, Ex = 0 ), то тест є легким. Його валідизація полягає в ускладненні тестових завдань.

5) Якщо асиметрія є додатною, а ексцес від’ємним (As> 0, Ex< 0), то тест є складним з великою розподільною здатністю. Його валідизація полягає у валідизації змісту шляхом вилучення тестових завдань, які не відповідають цій вимозі.

6) При додатних асиметрії та ексцесі (As> 0, Ex> 0) тест є складним з малою розподільною здатністю. Це є наслідком дії двох факторів, що впливають один на одного: порушення валідності змісту, невалідність тестових завдань. Валідизація тесту полягає у валідизації тестових завдань.

7) При додатній асиметрії та відсутності ексцесу (As> 0, Ex = 0 ) тест є складним. Це є наслідком дії двох факторів, що впливають один на одного: порушення валідності змісту, невалідність тестових завдань. Валідизація тесту полягає у валідизації тестових завдань. Особливої уваги потребують два останніх критерії (8 і 9), які дають можливість визначити конструктивну валідність тесту, а саме перевірити тест на однорідність. Нагадаємо, що тест вважається однорідним, якщо його сконструйовано з тестових завдань з близькими характеристиками. Неоднорідним вважається такий тест, який складається з окремих субтестів, кожний з яких можна розглядати як тест з певними характеристиками.

8) Якщо крива розподілу симетрична, асиметрія дорівнює нулю, а ексцес додатний (As = 0, Ex> 0), то тест є неоднорідним і сконструйованим із двох субтестів: складного і легкого. Валідизація тесту полягає в його поділі на два окремих тести.

9) Якщо крива розподілу симетрична, асиметрія дорівнює нулю, а ексцес від’ємний (As = 0, Ex< 0), то тест є неоднорідним, сконструйованим із трьох субтестів: легкого, середнього і складного. Валідизація тесту полягає в його поділі на окремі тести.

Застосуємо ці критерії до отриманих результатів тестування. В результаті тестування були отримані значення асиметрії та ексцесу за допомогою внутрішніх засобів Moodle. Ці характеристики наведені в таблиці 1.

*Таблиця 1 – Значення асиметрії та ексцесу, отримані в результаті тестування по дисциплінам кафедри*

|  |  |  |
| --- | --- | --- |
| Назва дисципліни | Значення асиметрії  розподілу(для найвище оціненої спроби) | Значення ексцесу  розподілу (для найвище оціненої спроби) |
| Обчислювальна техніка та програмування | -1,846 | 3,639 |
| Чисельні методи та моделювання на ЕОМ | -0,6541 | 0,4091 |
| Економічна інформатика | -0,8872 | 0,2748 |

Оскільки тести виявилися невалідними, то виникає необхідність провести більш ретельний аналіз з метою виявлення легких питань та питань з незадовільною розподільною здатністю. Це можливо зробити засобами системи Moodle.

Для реалізації цієї мети викладачами кафедри інформатики і інженерної графіки Донбаської державної машинобудівної академії за допомогою вбудованого в Moodle статистичного модуля були отримані наступні характеристики для тестових завдань: успішність, стандартне відхилення, оцінка навмання, призначена вага, ефективна вага, розрізнення, ефективність розрізнення.

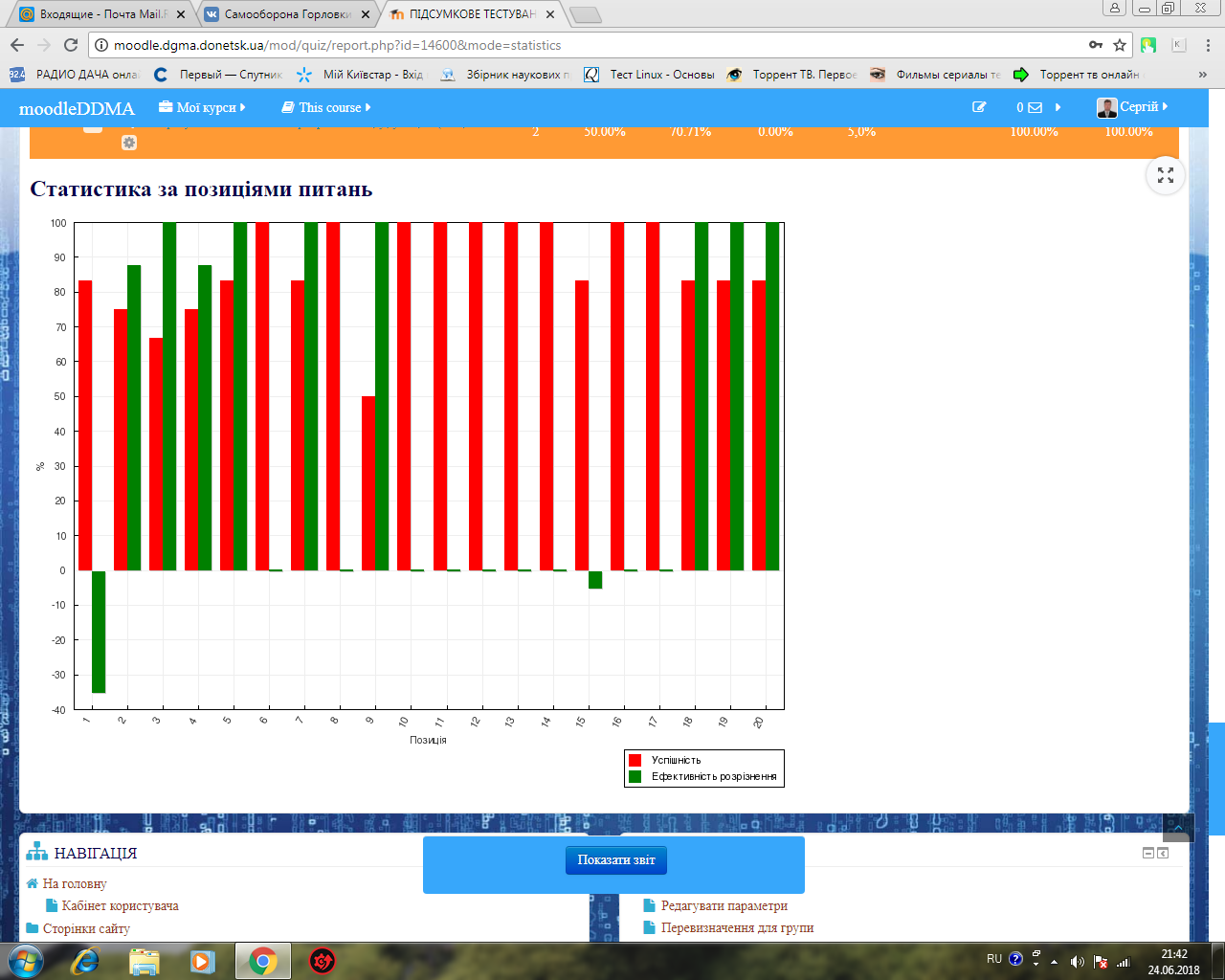
Параметр успішність  визначає ту частину студентів, яка правильно відповіла на поставлене запитання і обчислюється по формулі

,

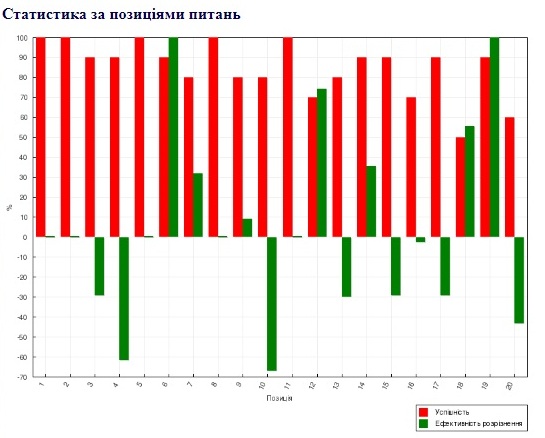
де *m*– кількість тестових завдань, *n*- кількість студентів, що брали участь у тестуванні, - оцінка успішності виконання *j*-го завдання *i*-м студентом. Чим більше величина коефіцієнта , тим більша кількість студентів виконала *j*-те запитання [6-9].

Складність завдань тесту мусить відповідати рівню підготовки студентів,що проходять тестування.Тест в цілому повинен включати в себе комплекс завдань різної складності – від легких до важких [1]. Занадто прості завдання, на які правильно відповідають всі випробувані () i занадто складні завдання, на які не може відповісти ніхто з опитуваних () не володіють здатністю диференціювати тестуємих за рівнем їх підготовки і повинні з виключатися з переліку тестових завдань.

На рисунках 1,2 і 3 наведені гістограми значень коефіцієнта успішності та ефективності розрізнення тестових завдань, які були отримані в результаті підсумкового тестування по дисциплінам курсу "Обчислювальна техніка та програмування" , "Чисельні методи та моделювання на ЕОМ" і "Економічна інформатика"

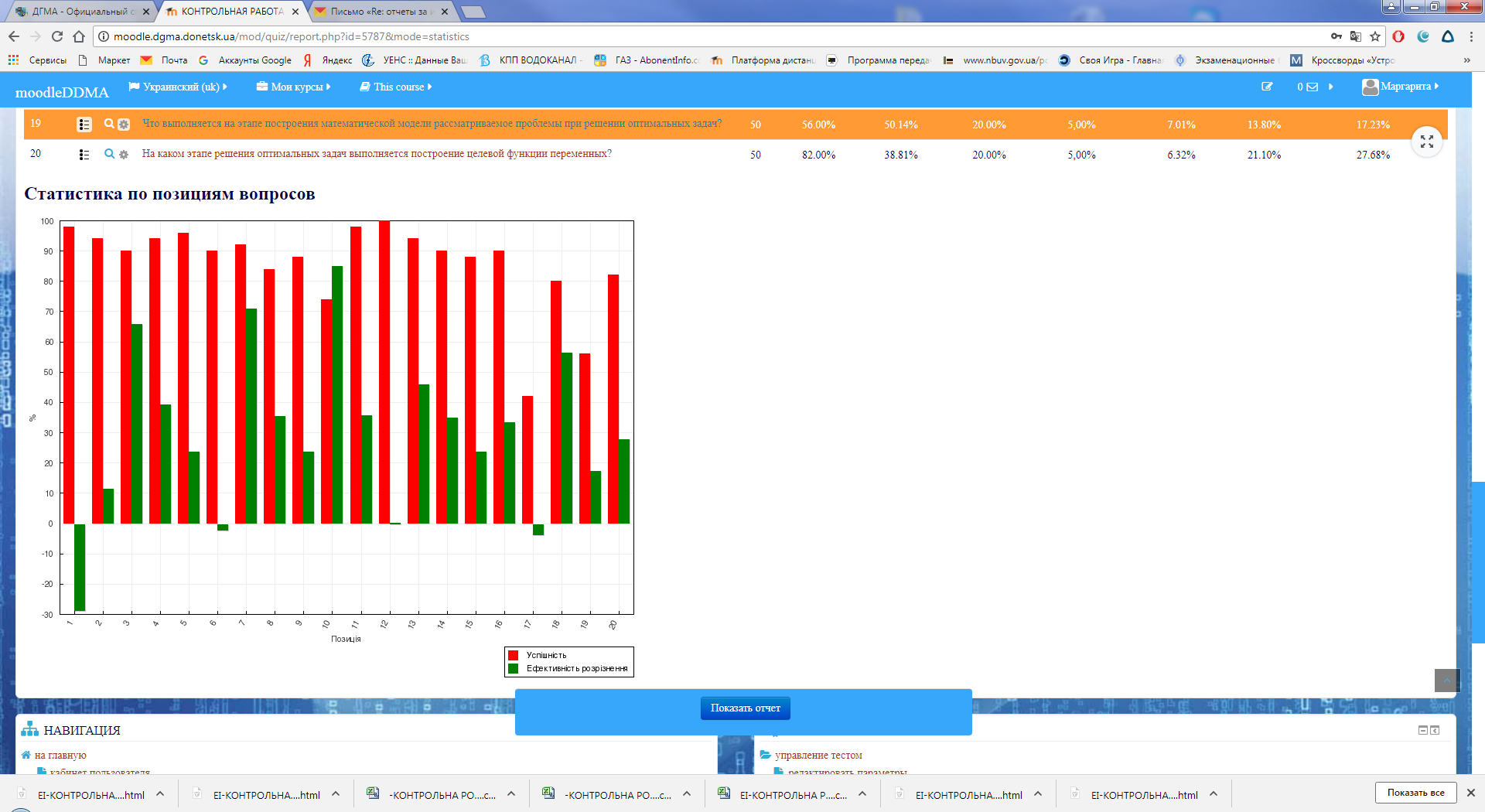


*Рисунок 1 – Значення коефіцієнтів успішності та ефективності розрізнення завдань підсумкового тесту по курсу "Обчислювальна техніка та програмування*"



*Рисунок 2 – Значення коефіцієнтів успішності та ефективності розрізнення завдань підсумкового тесту по курсу "Чисельні методи та моделювання на ЕОМ"*

Оцінка індексу успішності дозволила виявити в тесті неінформативні питання. Так, у тесті по дисципліні "Обчислювальна техніка та програмування" зустрічаються дуже легкі питання з номерами 6,10,11,12,13,14,16,17 на які відповіли всі опитувані. У тесті по дисципліні "Чисельні методи та моделювання на ЕОМ" дуже легкими виявилися питання з номерами 1,2,5,8,11. З графіку на рисунку 3 видно , що серед завдань по дисципліні "Економічна інформатика" дуже легкими є запитання 1,11 і 12.



*Рисунок 3 ­ Значення коефіцієнтів успішності та ефективності розрізнення завдань підсумкового тесту по курсу "Економічна інформатика"*

Стандартне відхилення результатів випробовуваних по *j* -му завданню тесту обчислюється за формулою:

,

де – оцінка успішності виконання *j*-го завдання *i*-м студентом, - середня оцінка, отримана *i*-м студентом за виконання тесту. Аналіз величин стандартного відхилення оцінки для кожного питання дозволяє виявити її внесок в розподільну здатність тесту. Для більшості тестових питань в досліджених тестових завданнях воно має значення більше 0,30, що, відповідно з вимогами педагогічної теорії вимірювань [1-3], є задовільним показником їх розподільної здатності (рисунок 3). Завдання, для яких це значення менше 0,30, не маютрозподільної здатності і повинні бути перероблені. Наприклад, у випадку тесту по дисципліні "Чисельні методи та моделювання на ЕОМ" завдання з номерами 1,2,5,11 мають стандартне відхилення, яке дорівнює нулю і тому їх треба видалити з тесту.

Введемо деякі позначення і визначимо їх наступним чином: *s* - номер спроби здачі тесту, *S* - множинавсіх спроб здачі тесту,  - кількість набраних студентом балів за тест, - оцінка студента за визначене завдання, - оцінка студента за всі інші завдання:





Дисперсія показника позначатиметься V(), наприклад:



Коваріація позначатиметься С(), наприклад:



Важливою статистичною характеристикою розподільної здатності тестових завдань, яку можна отримати засобами Moodle, є розрізнення [9]:



Цей коефіцієнт показує, наскільки взаємопов'язані правильність відповіді на дане питання і інші питання тесту. Для гарного тестового питання передбачається, що студенти з високими оцінками за нього також будуть мати більш високі оцінки і за тест в цілому. На це вказує позитивне значення коефіцієнта. Ефективність розрізнення  – це нормований коефіцієнт розрізнення[10].

Ефективність розрізнення є мірою здатності конкретного завдання розділяти сильних і слабих випробовуваних. Позитивні значення відповідають завданням, які дійсно розділяють «сильних» і «слабих» студентів, в той час як від'ємне значення  свідчить про те, що погано підготовлені студенти відповідають на дане завдання в середньому краще, ніж добре підготовлені. Очевидно, що такі завдання не є тестовими, тому що не здатні адекватно розділяти випробовуваних за рівнем їх підготовки, і їх слід вибраковувати. Вважається що завдання має достатню розподільну здатність, якщо ефективність розрізнення має значення більше або рівне 0,30 [1, 2].

Аналіз величини  для розглянутих тестів дозволив виявити завдання, що не мають достатньої розподільної здатності (0,30). Аналіз даних на рисунку 1 свідчить про те, що серед тестових питань по дисципліні "Обчислювальна техніка та програмування" завдання з номерами 1,15 мають від'ємну ефективність розрізнення і потребують переробки. Для тестових завдань по дисципліні " Чисельні методи та моделювання на ЕОМ" наприклад, питання з номерами 3,4,9,10,13,15 не вищенаведеним вимогам, крім того, для деяких з наведених питань значення ефективності розрізнення виявилося від'ємним, що свідчить про необхідність перегляду цих питань(рисунок 2). Щодо тестових завдань по дисципліні "Економічна інформатика" питання 1, 6, 17 мають від'ємне значення коефіцієнту ефективності розрізнення, а питання 2,5,9,15,19 мають значення цього коефіцієнту менше 30 відсотків. Тому ці питання слід переробити.

Таким чином, за допомогою статистичних характеристик, обчислених у системі Moodle були знайдені легкі тестові питання та питання з низькою розподільною здатністю. Переробка цих питань дозволить отримати валідні тестові завдання. Отримання валідних тестів буде темою подальших досліджень.

**Література:**

1. Ким, В.С.Тестирование учебных дострижений. / В.С. Ким. – Уссурийск: Издательство УГПИ, 2007. – 214 с.
2. Методичні рекомендації щодо розроблення електронних навчально- методичних матеріалів в системі Moodle DDMA для студентів всіх форм навчання. Бурлуцький С.В.- Краматорськ:ДДМА,2017. – 48 c.
3. Булах І. Є. Створюємо якісний тест: навч. посіб. / І.Є.Булах, М.Р. Мруга— К.: Майстер;клас, 2006 — 160 с.
4. Анастази А. Психологическое тестирование: Пер. с англ. в 2 кн. / А. Анастази – Кн. 1. – М.: Педагогика, 1982. – 316 с.
5. Булах І. Є. Теорія комп’ютерного тестування. / I.Є.Булах. — К.:ЦМКМОЗ України, 1994. — 59 с.
6. Протасова И.В. Методика анализа и повышения качества тестов в системе електронного обучения Moodle. / И. В. Протасова, А. П. Толстобров, И. А. Коржик // Вестник Воронежского гос.ун.-та: Серия системный анализ и информационные технологии. – 2014. - №3 – С.61-72.
7. Коржик, И.А. Тестовая система Moodle и качество тестовых заданий / И.А. Коржик, И.В. Протасова, А.П. Толстобров // Современные информационные технологии и ИТ-образование: сб. избранных трудов VII междунар. науч.-практич. конф. – М: ИНТУИТ.РУ, 2012. – C. 187–196.
8. Толстобров, А.П. Возможности анализа и повышения качества тестовых заданий при использовании сетевой системы управления обу-чением MOODLE / А.П. Толстобров, И.А. Коржик // Вестник Воронежского гос. ун- та. системный анализ и информационные технологии. –2008. –№ 2. –с. 100–106.
9. Коржик, И.А. Оценка качества тестов в системе электронного обучения MOODLE [Электронный ресурс] / И.А. Коржик, А.П. Толстобров. – режим доступа: http://www.infoco.ru/ mod/data/view.php?d=4&rid=114
10. Аванесов В.С. Понятие и методы математической теории педагогических измерений (Item Response Theory): статья третья./ В.С.Аванесов // Педагогические Измерения. – 2009. - №4 - С. 5.
11. Официальный сайт системы MOODLE [Электронной ресурс]. – Режим доступа: <http://www.moodle.org>