

Твір наукового характеру

"Метод аналізу достатності інформації специфікації вимог до програмного забезпечення (ПЗ) для визначення складності та якості програмного забезпечення на основі порівняльного аналізу онтологій"

Однією із основних вимог користувачів до сучасного ПЗ є якість. Під якістю розуміють таку характеристику ПЗ, що відображає ступінь його відповідності вимогам. Ще однією важливою характеристикою ПЗ є його складність. Структурна складність програм визначається кількістю взаємодіючих компонент, кількістю зв'язків між компонентами та складністю їх взаємодії. На сьогодні існує ряд моделей, що дозволяють розраховувати якість та складність ПЗ, однак багатозначність трактування цих характеристик ускладнює такі розрахунки. Більшість моделей базуються на використанні різних метрик ПЗ. При цьому актуальною є задача оцінювання достатності інформації щодо ПЗ (наприклад, можливості отримання достовірної інформації щодо показників для обчислення значень метрик), на основі якої здійснюється визначення характеристик якості та складності. Неповнота, неточність та спотворення такої інформації призводять, відповідно, до падіння достовірності оцінок якості та складності.

Найбільш використовуваними джерелами інформації щодо складності ПЗ є метрики: Чепіна, Мак-Клура, Кафура, Холстеда, Маккейба, Джилба. Щодо якості ПЗ, то найчастіше для її оцінки використовуються метрики зв'язності, зчеплення, очікуваної вартості розроблення, прогнозованої вартості перевірки якості, прогнозованої вартості розроблення, очікуваного загального часу розроблення ПЗ, прогнозованого часу етапу проектування. Однак усі ці метрики і моделі оцінки якості та складності на їх основі мають ряд недоліків, що не дозволяють досягти бажаної достовірності визначення якості та складності ПЗ. При розробленні ПЗ вже на початкових етапах життєвого циклу, зокрема на етапі проектування, необхідно визначитися з моделями, метриками та показниками, на основі яких у подальшому будуть здійснюватися оцінки якості та складності.

Розглянемо показник «кількість модулів», що використовується у таких метриках оцінки складності як: метрика Чепіна, Джилба (абсолютна), Мак-Клура і Кафура. Якщо значення цього показника буде неточним, то, відповідно, знизиться ступінь достовірності усіх метрик, що його використовують. Якщо ж значення цього показника буде відсутнім у специфікації вимог до ПЗ, то зникне можливість розрахунку усіх чотирьох метрик складності програмного проекту.

Метрики Чепіна та Кафура мають спільний показник «кількість модулів», тому є взаємно корелюючими. Метрика Чепіна: $I_s = Q_m \cdot (P + 2M + 3C + 0,5T)$, де I_s - інформаційна міцність, Q_m - кількість модулів, P - середня кількість змінних для розрахунків і виведення в модулі; M - середня кількість модифікованих або створених в модулі змінних; C - середня кількість керуючих змінних модуля; T - середня кількість не використовуваних в модулі ("паразитних") змінних. Метрика Кафура розраховується як: $I = Q_m \cdot (W \cdot R + W \cdot W_{rRd} + W_{rRd} \cdot R + W_{rRd} \cdot (W_{rRd} - 1))$, де I - інформаційна складність, Q_m - кількість модулів, W - середня кількість процедур модуля, які оновлюють структуру даних, R - середня кількість процедур модуля, які читають інформацію зі структури даних, W_{rRd} - середня кількість процедур модуля, які читають і оновлюють структуру даних.

Кількість модулів програми, виходячи з наведених формул метрик Чепіна та Кафура, можна визначити як: $Q_m = \frac{I_s}{(P + 2M + 3C + 0,5T)}$ - на основі метрики Чепіна;

$Q_m = \frac{I}{W \cdot R + W \cdot W_{rRd} + W_{rRd} \cdot R + W_{rRd} \cdot (W_{rRd} - 1)}$ - на основі метрики Кафура.

Тоді: $\frac{I_s}{(P + 2M + 3C + 0,5T)} = \frac{I}{(W \cdot R + W \cdot W_{rRd} + W_{rRd} \cdot R + W_{rRd} \cdot (W_{rRd} - 1))}$. Отже,

метрику Кафура можливо обчислити з використанням метрики Чепіна (і навпаки):

$I = \frac{I_s \cdot (W \cdot R + W \cdot W_{rRd} + W_{rRd} \cdot R + W_{rRd} \cdot (W_{rRd} - 1))}{(P + 2M + 3C + 0,5T)}$, що свідчить про їх взаємну

кореляцію.

Якщо деякі з показників, що входять до складу цих метрик, визначені неточно або відсутні, то використання цих двох метрик одночасно у моделях чи комплексних показниках оцінки складності суттєво вплине на достовірність отриманих оцінок. За такої ситуації важливою є умова пом'якшення впливу взаємної кореляції таких метрик при використанні їх у моделях чи комплексних показниках оцінки складності та якості. Таке пом'якшення здійснюється шляхом виявлення спільних показників, підвищення точності їх значень, або, при можливості, обмеження одночасного задіювання наборів метрик, що містять однакові показники.

У цьому контексті значну цінність мають знання фахівців, що вже володіють досвідом оцінки якості та складності для різних типів ПЗ. Наприклад, цінними є знання щодо взаємовпливу та кореляції показників, котрі використовуються при обчисленні метрик та характеризують властивості ПЗ, оскільки така кореляція може погіршувати точність отриманих оцінок, частина показників, важливих для деякого типу ПЗ може упускатись взагалі, достовірність результуючих оцінок може значно знижуватись і т.і. Інформацію щодо визначення показників якості та складності ПЗ, наприклад, взаємозв'язків метрик за показниками, зручно подавати у вигляді семантичних мереж чи інших структур, що дозволяють відобразити причинно-наслідкові зв'язки між поняттями. Однією з таких структур є онтологія. Перевагами використання онтологій є системний підхід до вивчення предметної галузі, можливість цілісного подання відомої інформації предметної галузі, виявлення дублювань та прогалін у знаннях на основі візуалізації відсутніх логічних зв'язків.

Для побудови та візуалізації онтологій на сьогодні розроблено ряд програмних продуктів, зокрема і універсальних, що дають змогу працювати з різними предметними галузями: Ontolingua Server, SMART, Protégé, OntoEdit, WebOnto, ODE (Ontological Design Environment), DOE (Differential Ontology Editor), CONE, OntoEditor+. Для даної роботи використовувалось вільно поширюване програмне забезпечення Protégé 4.2, що розроблене групою медичної інформатики Стенфордського університету (перша версія – у 1987 році) з метою допомогти розробникам ПЗ у створенні та підтримці явних моделей предметної галузі та включення цих моделей безпосередньо у програмний код, але також дає можливість працювати (створювати, редагувати, переглядати та порівнювати) з онтологіями різних предметних галузей.

Для оцінки достатності інформації при визначенні складності та якості ПЗ пропонується метод на основі порівняльного аналізу онтологій.

Насамперед, для предметної галузі «Інженерія ПЗ» будується базова онтологія, у котрій є інформація щодо метрик, що застосовуються для визначення складності та якості ПЗ.

Для побудови такої онтології оберемо 4 характеристики ПЗ: складність програмного проекту, складність ПЗ, якість програмного проекту, якість ПЗ. Зазначені характеристики визначаються на основі метрик та моделей – для визначення складності програмного проекту використовуються 4 метрики, для

визначення складності ПЗ – 6 метрик, для визначення якості проекту – 5 метрик, для визначення якості ПЗ – 9 метрик. Відомі формули для визначення метрик та моделей дають можливість обрати показники специфікації та ПЗ для визначення відповідних метрик та моделей – для визначення метрик складності проекту необхідно 13 показників, для визначення метрик складності ПЗ – 10 показників, для визначення метрик якості проекту – 14 показників, для визначення метрик якості ПЗ – 13 показників.

Складовими частинами базової онтології є: базова онтологія для визначення складності програмного проекту (рис.1), базова онтологія для визначення складності ПЗ (рис.3), базова онтологія для визначення якості проекту (рис.5), базова онтологія для визначення якості ПЗ (рис.7).

Побудована базова онтологія дала можливість зробити наступні висновки:

1) всі чотири метрики для визначення складності програмного проекту використовують показник «кількість модулів»;

2) чотири з шести метрик (метрики Холстеда, Маккейба, Джилба та кількість операторів програми) для визначення складності ПЗ використовують показник «загальна кількість операторів», три (LOC-оцінка, метрика Холстеда та кількість операторів програми) з шести зазначених метрик використовують показник «кількість рядків коду», дві (метрика Холстеда та прогнозована оцінка складності інтерфейсів) з шести метрик використовують показник «загальна кількість операндів» та «кількість унікальних операндів»;

3) метрики для визначення якості проекту не мають спільних показників;

4) три з дев'яти метрик для визначення якості ПЗ (вартість проектування, вартість перевірки якості, вартість розроблення) використовують показник «ціна одного рядка коду», дві (трудовитрати та тривалість проекту за моделлю Боема) з дев'яти зазначених метрик використовують показник «тип проекту», вісім (крім функційного розміру) з дев'яти метрик використовують показник «кількість рядків коду», три (загальний час розроблення ПЗ, час етапу проектування, продуктивність розроблення ПЗ) з дев'яти метрик використовують показник «тривалість проекту»;

5) характеристики «складність проекту», «складність ПЗ» та «якість проекту» використовують показник «кількість модулів»; характеристики «складність проекту» та «якість проекту» використовують показник «кількості модулів, безпосередньо передуючих і безпосередньо слідуючих за модулем»; характеристики «якість проекту», «складність ПЗ» та «якість ПЗ» використовують показник «кількість рядків коду»; характеристики «якість проекту» та «якість ПЗ» використовують показник «прогнозована тривалість проекту», «частка (доля) етапу проектування у життєвому циклі».

Отже, при оцінюванні складності та якості ПЗ необхідно приділити особливу увагу тим вищевказаним показникам, що входять до складу декількох метрик одночасно.

Метод аналізу достатності інформації специфікації вимог до програмного забезпечення (ПЗ) для визначення складності та якості програмного забезпечення на основі порівняльного аналізу онтологій складається з наступних кроків:

1) аналіз специфікації вимог до ПЗ на предмет наявності показників, необхідних для визначення складності та якості програмного проекту, прогнозованої складності та якості ПЗ;

2) розроблення онтології для визначення складності та якості конкретного програмного забезпечення, складовими частинами якої є: онтологія для визначення складності конкретного програмного проекту, онтологія для визначення складності

конкретного ПЗ, онтологія для визначення якості конкретного проекту, онтологія для визначення якості конкретного ПЗ;

3) порівняння розробленої онтології з базовою онтологією для визначення складності та якості ПЗ, складові частини якої наведені на рис.1, 3, 5, 7;

4) виявлення показників, які відсутні в онтології для визначення складності та якості конкретного програмного забезпечення;

5) виявлення метрик, які неможливо обчислити на основі наявних показників;

6) виявлення характеристик, значення яких неможливо визначити на основі метрик, які залишається можливим обчислити при наявній інформації;

7) наявність характеристик, значення яких неможливо визначити на основі специфікації вимог до ПЗ, свідчить про вимогу доповнення цієї специфікації необхідними для обчислення тієї чи іншої характеристики показниками (на цьому етапі можливе як додавання необхідної інформації, так і видалення іншої інформації);

8) повторення етапів 2-7 доти, доки не стане можливим визначити всі необхідні характеристики або доки не буде сформовано висновок, що недостатньо даних для визначення складності та якості ПЗ.

При проведенні дослідження було проаналізовано специфікацію вимог до ПЗ автоматизованої системи широкоформатного фотодруку. На основі аналізу специфікації визначено показники, необхідні для визначення складності та якості програмного проекту, прогнозованої складності та якості ПЗ. Ці показники дали змогу розробити онтологію для визначення складності та якості даного програмного забезпечення, складовими частинами якої є: онтологія для визначення складності даного програмного проекту (рис.2), онтологія для визначення складності даного ПЗ (рис.4), онтологія для визначення якості даного проекту (рис.6), онтологія для визначення якості даного ПЗ (рис.8).

Порівняння розробленої онтології для програмного проекту з розроблення автоматизованої системи широкоформатного фотодруку з фрагментами базової онтології для предметної галузі «Інженерія ПЗ» дало можливість виявити, що у розробленій онтології відсутні 9 показників: кількість керуючих змінних, ціна одного рядка коду, тривалість програмного проекту, тип програмного проекту (за Боемом), кількість рядків коду, кількість зв'язків кожного модуля, кількість модулів, частка (доля) етапу проектування у життєвому циклі, загальна кількість операторів – показники специфікації вимог до ПЗ.

Крім цього, на основі порівняння розробленої онтології для програмного проекту з розроблення автоматизованої системи широкоформатного фотодруку з базовою онтологією було виявлено, що у розробленій онтології недостатньо даних для визначення деяких метрик та моделей через відсутність вищевказаних показників.

Аналіз рис.1 та рис.2 дав можливість зробити висновок, що для обчислення метрик Чепіна, Мак-Клура та Кафура недостатньо даних, а для обчислення метрики Джилба (абсолютна складність) взагалі відсутні необхідні показники, отже, жодну з метрик складності проекту не можна обчислити, тому відповідно не можна визначити складність програмного проекту.

Аналіз рис.3 та рис.4 показав, що для обчислення метрик Холстеда, Маккейба, Джилба (відносна складність) та прогнозованої оцінки складності інтерфейсів ПЗ наявних показників недостатньо, а для обчислення LOC-оцінки та прогнозованої кількості операторів програми необхідні показники взагалі відсутні, отже, жодну з метрик прогнозування складності ПЗ не можна обчислити, тому не можна визначити й прогнозовану складність розроблюваного ПЗ.

Аналіз рис.5 та рис.6 дав можливість зробити висновок, що для обчислення часу модифікації моделей необхідні показники відсутні, для обчислення кількості

знайдених помилок при інспектуванні моделей наявної інформації недостатньо, а метрики зв'язності, зчеплення та звертання до глобальних змінних можуть бути обчислені. Отже, якість програмного проекту можна визначити лише на основі метрик зв'язності, зчеплення та звертання до глобальних змінних (три метрики з п'яти – недостатність інформації).

Аналіз рис.7 та рис.8 показав, що для обчислення загального часу розроблення ПЗ, часу виконання робіт процесу проектування, очікуваної вартості розроблення ПЗ, продуктивності розроблення ПЗ, оцінок трудовитрат та тривалості проекту за моделлю Боєма необхідні показники відсутні, а для обчислення вартості перевірки якості та витрат на реалізацію програмного коду наявної інформації недостатньо. Обчислена може бути лише одна метрика – прогнозований функційний розмір, тому прогнозовану якість ПЗ можна визначити лише на основі однієї метрики з дев'яти (недостатність інформації).

Тоді відсутність у специфікації вимог до ПЗ дев'яти показників призвела до неможливості обчислення 20 метрик (з 24) та відповідно до неможливості визначення складності проекту та розроблюваного ПЗ і до неповноти визначення якості проекту та розроблюваного ПЗ.

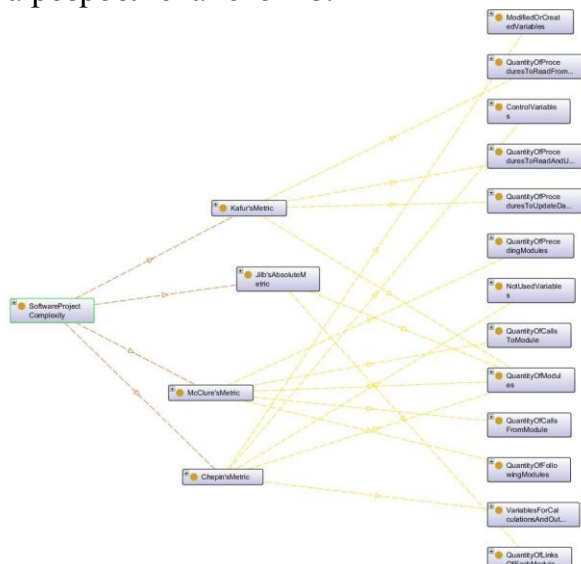


Рис. 1. Базова онтологія для визначення складності програмного проекту

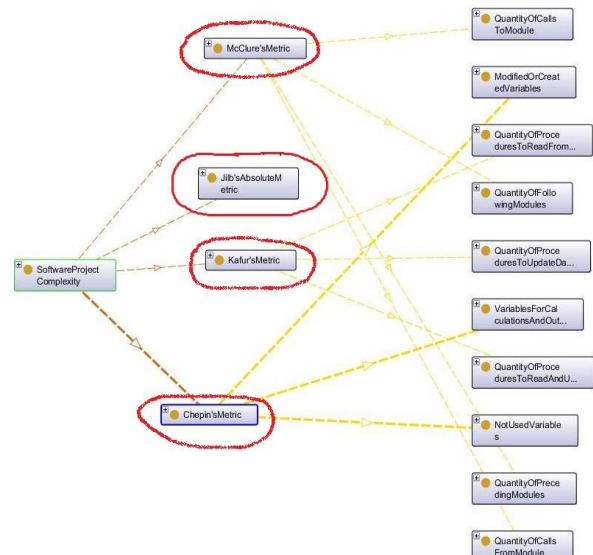


Рис. 2. Онтологія для визначення складності програмного проекту з розроблення автоматизованої системи (АС) широкоформатного фотодруку

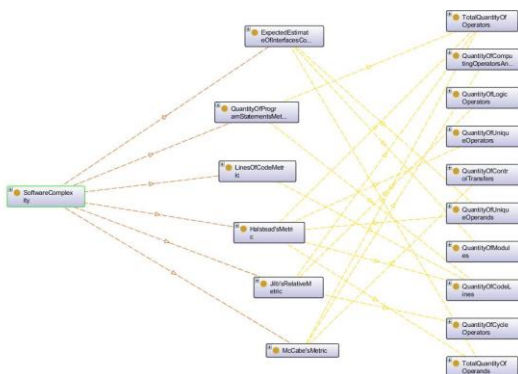


Рис. 3. Базова онтологія для прогнозування складності ПЗ

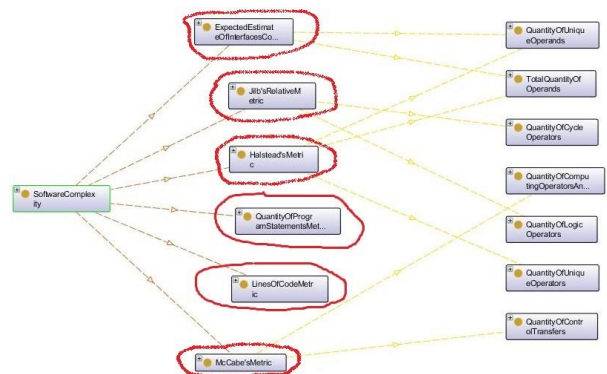


Рис. 4. Онтологія для прогнозування складності ПЗ АС широкоформатного фотодруку

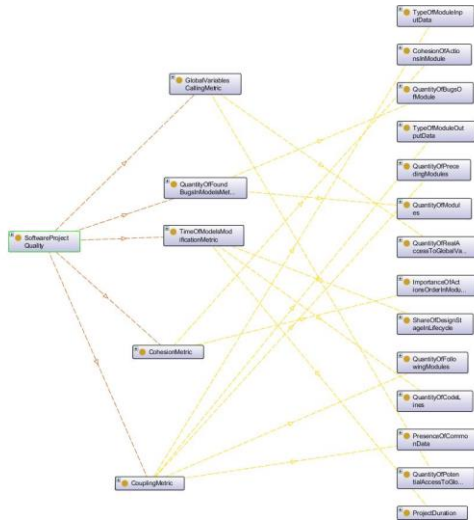


Рис. 5. Базова онтологія для визначення якості програмного проекту

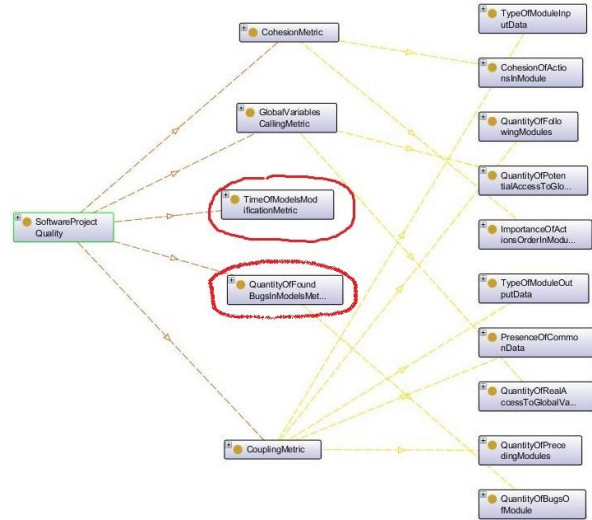


Рис. 6. Онтологія для визначення якості програмного проекту з розроблення АС широкоформатного фотодруку

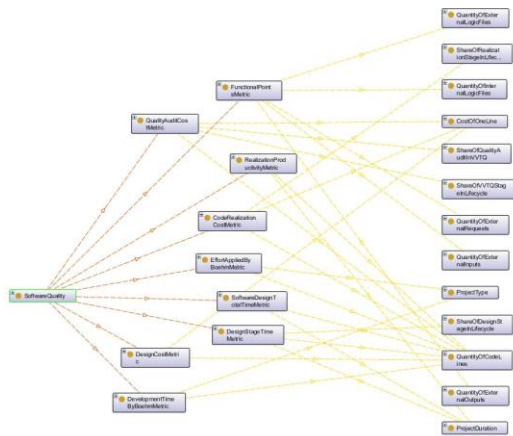


Рис. 7. Базова онтологія для прогнозування якості ПЗ

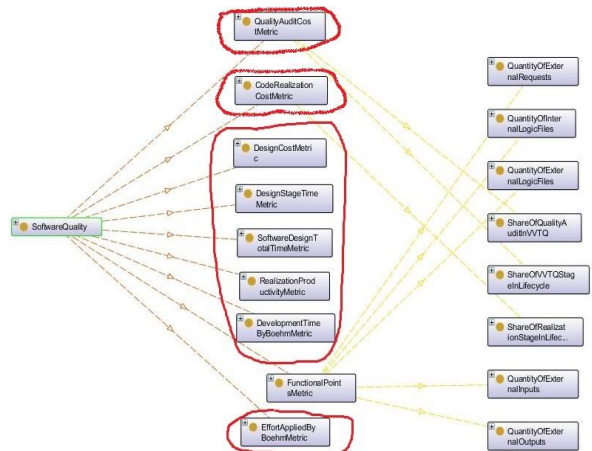


Рис. 8. Онтологія для прогнозування якості ПЗ АС широкоформатного фотодруку

Оскільки для даного програмного проекту є характеристики, які при наявній в специфікації вимог інформації взагалі неможливо визначити або можна визначити в недостатній мірі, то відбулось доповнення специфікації вимог до ПЗ, після чого було створено нову версію онтології для визначення складності та якості конкретного програмного забезпечення.

Порівняльний аналіз нової версії онтології із базовою онтологією показав, що зміни відбулись при визначенні складності конкретного програмного проекту (рис.9), при прогнозуванні складності конкретного ПЗ (рис.10) та при визначенні якості проекту (рис.11)

Порівняння нової версії онтології для визначення складності та якості конкретного програмного забезпечення та базової онтології показало, що у розробленій онтології відсутні 14 показників, тобто при доповненні специфікації вимог до ПЗ було додано лише 2 показники: кількість модулів та загальна кількість операторів. Але після додавання цих показників з'явилась можливість обчислити: метрики Мак-Клура і Кафура, та визначити складність програмного проекту на основі двох метрик (з чотирьох) – рис.9; метрики Маккейба, Джилба (відносна складність), прогнозовану оцінку складності інтерфейсів, що дає можливість визначити прогнозовану складність ПЗ на основі трьох метрик (з шести) – рис.10; кількість знайдених помилок при інспектуванні моделей, що дає можливість визначити якість проекту на основі вже чотирьох з п'яти метрик – рис.11. Отже, доповнення

специфікації вимог до ПЗ двома показниками збільшило кількість інформації для визначення якості проекту та дало можливість визначити складність проекту та розроблюваного ПЗ (все ще недостатність інформації).

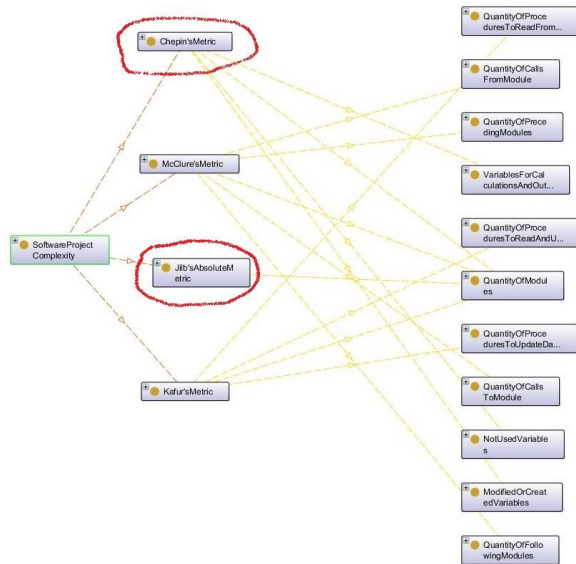


Рис. 9. Онтологія (версія 2) для визначення складності програмного проекту з розроблення АС широкоформатного фотодруку

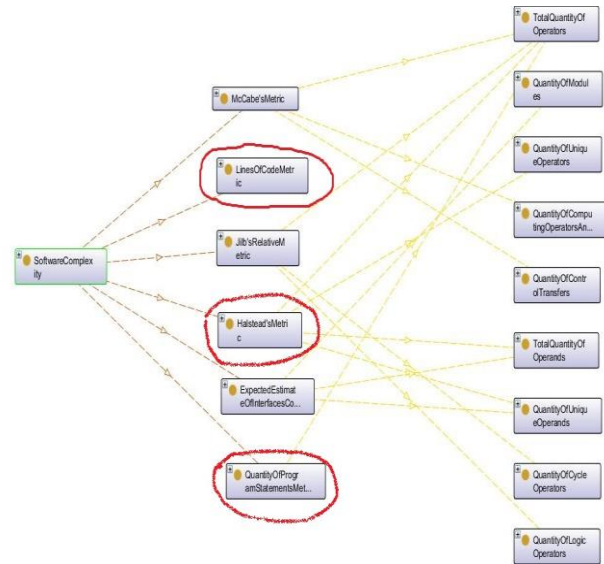


Рис. 10. Онтологія (версія 2) для прогнозування складності ПЗ АС широкоформатного фотодруку

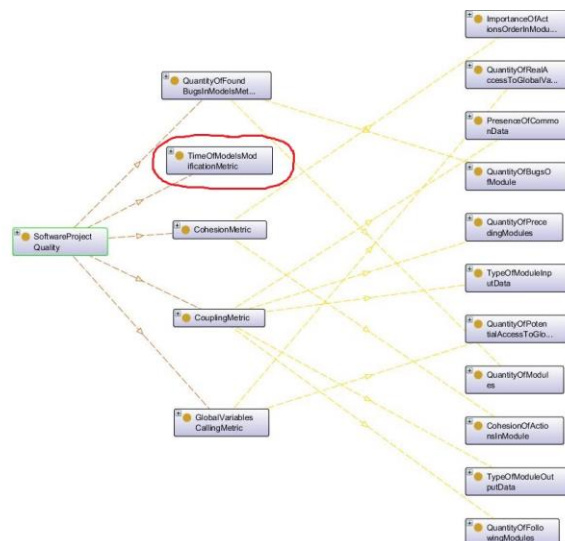


Рис. 11. Онтологія (версія 2) для визначення якості програмного проекту з розроблення автоматизованої системи широкоформатного фотодруку

На основі методу оцінки достатності інформації для визначення складності та якості ПЗ і аналізу базових онтологій можна зробити наступні рекомендації щодо формування специфікації вимог до ПЗ (з точки зору можливості оцінювання зазначених характеристик):

1) за потреби визначати складність проекту за чотирма метриками, специфікація вимог до ПЗ повинна містити наступні 13 показників: кількість змінних для розрахунків і виведення, кількість модифікованих або створених в модулі змінних, кількість керуючих змінних, кількість невикористовуваних в модулі змінних, кількість модулів, кількість зв'язків кожного модуля, кількості модулів, безпосередньо передуючих і безпосередньо слідуючих за модулем, кількість звертань до модуля, кількість викликів з модуля, кількість процедур, які оновлюють структуру даних, кількість процедур, які читають інформацію зі структури даних, кількість процедур, які читають і оновлюють структуру даних;

2) за потреби прогнозувати складність ПЗ за шістьма метриками, специфікація

вимог до ПЗ повинна містити наступні 10 показників: прогнозована кількість рядків коду, прогнозована кількість унікальних операторів, прогнозована кількість унікальних операндів, прогнозована загальна кількість операторів, прогнозована загальна кількість операндів, прогнозована кількість обчислювальних операторів та виразів, прогнозована кількість передач керування, прогнозована кількість логічних операторів, прогнозована кількість циклічних операторів, кількість модулів;

3) за потреби визначати якість проекту за п'ятьма метриками, специфікація вимог до ПЗ повинна містити наступні 14 показників: зв'язність дій всередині модуля, важливість порядку дій всередині модуля, типи вхідних та вихідних параметрів модуля, наявність спільних даних, кількості модулів, безпосередньо передуючих і безпосередньо слідуючих за модулем, кількість реальних викликів глобальної змінної, кількість можливих викликів глобальної змінної, прогнозована кількість рядків коду, прогнозована тривалість проекту, частка (доля) етапу проектування у життєвому циклі, кількість помилок в модулі, кількість модулів;

4) за потреби прогнозувати якість ПЗ за дев'ятьма метриками, специфікація вимог до ПЗ повинна містити наступні 13 показників: прогнозована кількість рядків коду, прогнозована тривалість проекту, частка (доля) етапу проектування у життєвому циклі, ціна одного рядка коду, частка (доля) етапу тестування, верифікації, валідації та аудиту якості у життєвому циклі, частка (доля) аудиту якості у етапі тестування, верифікації, валідації та аудиту якості, частка (доля) етапу реалізації у життєвому циклі, кількість зовнішніх входів, кількість зовнішніх виходів, кількість зовнішніх запитів, кількість внутрішніх логічних файлів, кількість зовнішніх логічних файлів, тип проекту (за Боемом).

В процесі аналізу метрик ПЗ, як джерел інформації щодо його характеристик, було виявлено наявність взаємної кореляції таких метрик, оскільки вони мають ряд спільних показників. Кореляція метрик може погіршувати точність та достовірність визначення характеристик ПЗ.

Знання досвідчених фахівців щодо взаємовпливу та кореляції метрик, котрі характеризують властивості ПЗ є цінними. Такі знання варто зберігати та використовувати при оцінюванні системних специфікацій ПЗ, що розробляється вперше, на предмет достатності інформації для визначення характеристик ПЗ.

У якості засобу відображення цих знань обрано онтології, які стали основою методу оцінки достатності інформації для визначення складності та якості програмного забезпечення.

Запропонований метод дав можливість розробити рекомендації щодо покращення специфікації вимог до ПЗ, що проілюстровано на прикладі визначення показників якості та складності ПЗ АС широкоформатного фотодруку.