

ОЦЕНКА ПОКАЗАТЕЛЕЙ РОСТА СЛОЖНОСТИ ЗАДАЧ АНАЛИЗА И КОММЕНТИРОВАНИЯ ПРОГРАММНЫХ КОДОВ

Пустыгин А.Н. (г.Челябинск)

Использование исходных текстов программ при анализе качества ПО сдерживается отсутствием средств автоматизации, поскольку ознакомление с большими объемами текстов с помощью текстового редактора или даже интегрированной среды разработки происходит очень медленно. Это положение вещей плохо соответствует наличию больших массивов программ с открытым исходным текстом, хранящих овеществленные знания поколений алгоритмистов-разработчиков ПО.

В больших программных проектах, созданных многочисленными коллективами программистов, комментарии к отдельным программным модулям могут быть неполными, недостоверными или просто отсутствовать. При внесении изменений программного обеспечения в процессе эксплуатации, даже содержащих комментарии, трудно ожидать их согласования с комментариями, оставшихся от предшествующих версий программного обеспечения.

Вторая сторона проблемы — использование исходного текста ПО для целей сопровождения программных продуктов. Реалии службы поддержки относятся к ряду коммерческих сведений, не подлежащих разглашению, однако технология сопровождения ПО сегодня далека от совершенства.

"Изучение" программы специалистом службы технической поддержки - это получение знаний об устройстве и принципах функционирования программного продукта, которые необходимы для квалифицированного сопровождения, то есть ответов на запросы пользователей и внесения исправлений при возникновении проблем эксплуатации.

Как известно, синтаксис языков программирования допускает комментарии двух видов — авторские и специализированные [1], последние могут генерироваться средой программирования автоматически. Как правило, авторский комментарий является текстом на естественном языке, но может содержать также элементы какой-либо формальной нотации, например математической. Важно заметить, что источником возникновения авторских комментариев является программист. Однако синтаксис комментария позволяет вводить метаописания, являющиеся эквивалентным описанием исходного кода на каком-либо метаязыке структурированного описания текста, например XML [2]. Такого рода комментарии предлагается называть техническими.

На сегодняшний день уже известно несколько сотен языков программирования и десятки их модификаций, и разработка новых языков программирования постоянно продолжается. Точные сведения о числе естественных языков и языков программирования привести сложно. Имеются лишь приблизительные оценки. Число естественных языков составляет от 6000 до 7000 [3], число языков программирования с вариантами по разным источникам — от 2500 до 3000 [4,5,6]. Есть оценка в 8512! [7,8].

Оценить характер динамики количества и состава языков программирования сложно, еще сложнее дать такую оценку для естественных языков. Однако в настоящий момент понятно, что количества естественных языков и языков программирования уже сравнимы, и если количество живых естественных

языков имеет тенденцию к сокращению [3], то тенденция развития языков программирования пока состоит в устойчивом росте. По материалам [9] эта зависимость похожа на линейную, хотя критерии отбора языков программирования при анализе явно не названы.

Рост объема ПО со временем обозначается термином «Раздувание программного обеспечения» [10,12]. Как правило, требования заказчика в части допустимого объема программ не принимаются во внимание, во всяком случае для ПО широкого применения. Более того, вопрос потребляемых аппаратных ресурсов разработчики массового ПО и изготовители вычислительной техники массового применения решают без учета мнения будущего потребителя.

Логарифмируя экспоненциальный закон роста данных по времени $N=N_0 \exp(\alpha * t)$, по соотношению $\lg N = \lg N_0 + \alpha * t / \lg e$ построим в логарифмическом масштабе кривые роста (рис. 1):

- числа слов в википедии [13],
- число оригинальных сигнатур вредоносных программ по данным Лаборатории Касперского [14],
- число оригинальных сигнатур вирусов по данным компании Semantec [15],
- объем потребляемой оперативной памяти Microsoft Windows [11],
- объем информации в internet [16],
- число строк ядра linux & freebsd [17,18,19].

Если экстраполировать полученные прямые в диапазонах, дополняющих до интервала 1995-2012 г.г. (тонкие линии на рисунке), можно заключить, что темп роста объема ПО на модели сигнатур вирусов (по двум независимым источникам) совпадает и хорошо согласуется с ростом содержания википедии. Объем потребляемой оперативной памяти Microsoft Windows растет медленнее, а скорость роста близка к скорости роста оценки информационной ёмкости всемирной паутины. Зависимость роста исходных текстов ядра операционных систем с открытым исходным кодом тоже экспоненциальна, однако отстает от первых двух групп. Главным выводом из представленных зависимостей является общая тенденция экспоненциального роста объемов информации как на естественных языках, так и на языках программирования. Это обстоятельство не оставляет никаких сомнений в необходимости автоматизации процесса извлечения знаний из исходных текстов, в частности на языках программирования, для противодействия порабощению человечества информацией, которую не представляется возможным воспринимать.

Если принять за базу объем трудозатрат для изучения исходного текста Windows95, то объем трудозатрат по решению аналогичной задачи для NT2000 увеличивается в 8 раз, а для Windows 7 — в 125 раз, что делает такую задачу практически не решаемой без инструментальных средств исследования исходных текстов. Если предположить, что характер рассматриваемых зависимостей не претерпит качественных изменений, можно спрогнозировать контрольные цифры для каждой из величин к 2030 году, и эти величины весьма велики (см. таблицу).

Прогнозируемый переход к автоматическому генерированию текстов ПО означает увеличение объема исходных текстов на несколько порядков, что приведет к информационному коллапсу систем программирования, и только своевременная разработка инструментария автоматического комментирования оставляет шанс предотвратить подобное развитие событий. Для каждого из известных искусственных языков программирования имеется некоторое множество (иногда весьма многочисленное) программ, и каждая из программ должна содержать комментарии её разработчика.

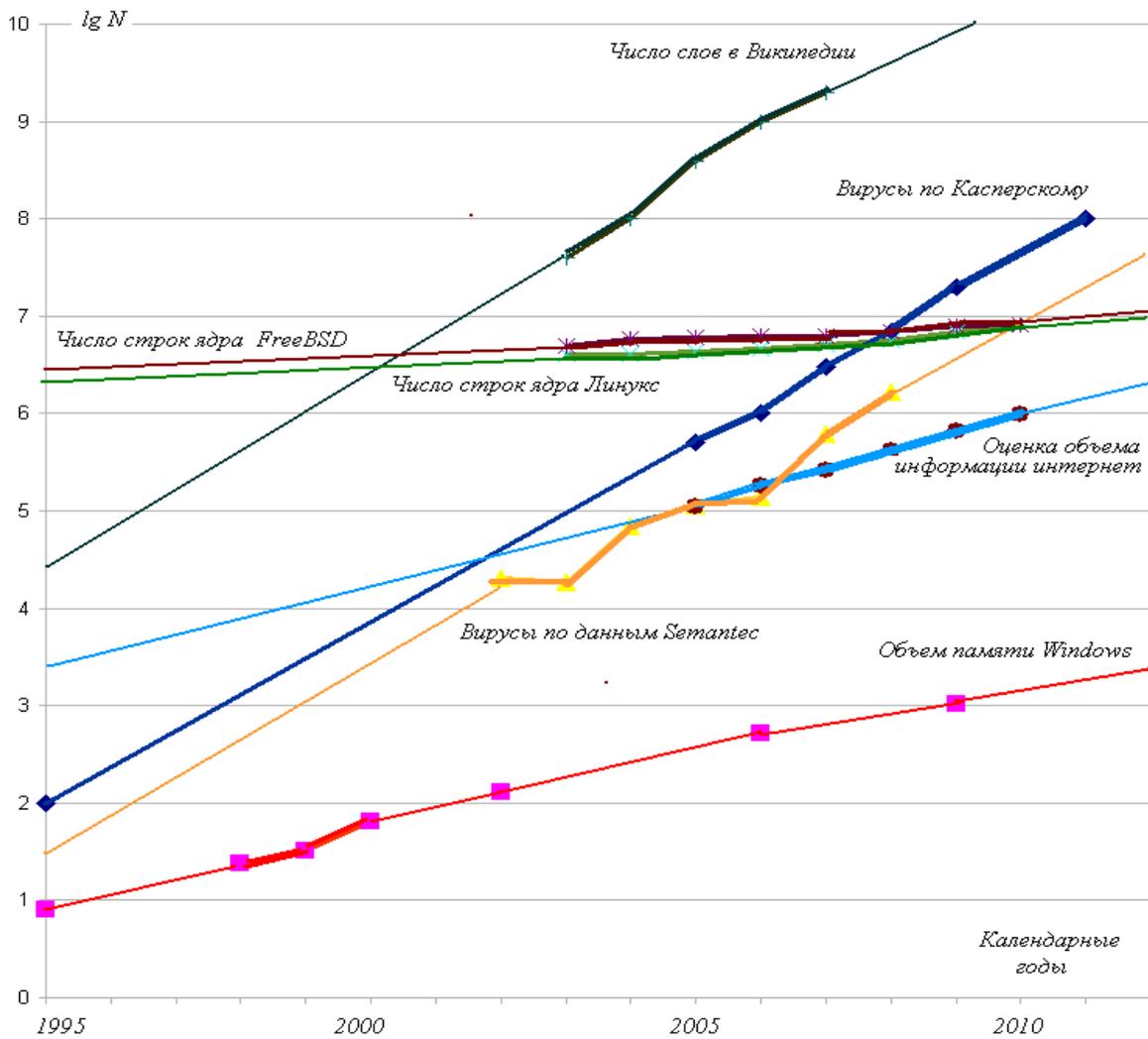


Рис. 1. Зависимости роста объемов текстов на естественных языках и языках программирования по данным публичных источников

Числовые данные рисунка 1

годы	$lg N$, где N - количество элементов данных						
1995		0,90					
1998		1,38					
1999		1,50					
2000		1,80					
2002		2,10	4,31				
2003			4,27	6,60	6,69		7,60
2004			4,83	6,60	6,77		8,00
2005	5,69		5,05	6,63	6,77	5,04	8,59
2006	6,00	2,70	5,14	6,67	6,79	5,25	9,00
2007	6,47		5,79	6,71	6,79	5,41	9,30
2008	6,84		6,22	6,75	6,84	5,61	
2009	7,30	3,01		6,83	6,90	5,81	
2010				6,91	6,90	5,99	
2011	8,00						
Прогноз 2030	15,71	6,14	13,2	7,79	7,71	9,811	19,075
	Число вирусов Касперский	Объем ОЗУ Windows	Число вирусов Semantec	Ядро Линукс	Ядро FreeBSD	Объем интернет	Объем Википедии

По трудозатратам на комментирование уже отлаженных программ допустимо тратить до 10% от общих трудозатрат. То есть эквивалентная стоимость машинного перевода исходных текстов в комментарии к ним оказывается значительной и в случае полного отсутствия комментариев она может составить от 10% до 0,3% от стоимости самого программного обеспечения. Информационный поиск показывает, что на данный момент средств автоматического (автоматизированного) синтеза комментариев к имеющемуся программному обеспечению нет. Это свидетельствует о некоторой неполноте применяемых систем программирования. Таким образом, одной из востребованных практикой задач является машинный синтез комментариев к текстам программ на исходном языке.

Литература:

1. Энциклопедия "SEO: Поисковая Оптимизация от А до Я" Бесплатный учебник по поисковой оптимизации и раскрутке сайтов. Комментарии (программирование) [http://www.seobuilding.ru/wiki/Комментарии_\(программирование\)](http://www.seobuilding.ru/wiki/Комментарии_(программирование)) — 2010
2. Расширяемый язык разметки XML. <http://ru.wikipedia.org/wiki/XML> — 2011
3. А.Е. Кибрик База естественного человеческого языка и ее основные параметры <http://www.dialog-21.ru/dialog2011/materials/html/3.htm> — 2011
4. Antula: Профессиональная студия веб-дизайна <http://www.antula.ru/power-language.htm> -2011
5. В. Kinnersley Collected Information On About 2500 Computer Languages, Past and Present. <http://people.ku.edu/~nkinners/LangList/Extras/langlist.htm> — 2011
6. Языки программирования <http://www.ikiznanie.ru/ru-wz/index.php/> — 2011
7. Язык программирования http://ru.wikipedia.org/wiki/Язык_программирования
8. И.Н.Бекман КОМПЬЮТЕРНЫЕ НАУКИ. Курс лекций. Лекция 8. ЯЗЫКИ ПРОГРАММИРОВАНИЯ - [profbeckman.narod.ru>Komp.files/Lec8.pdf](http://profbeckman.narod.ru/Komp.files/Lec8.pdf) —2011
- 9.Хронология_языков_программирования <http://ru.wikipedia.org/wiki/> — 2011
10. E. Perratore et al., "Fighting Fatware," Byte, Vol. 18 No. 4 Apr. 1993, pp. 98-108.
11. http://ru.wikipedia.org/wiki/Раздувание_программного_обеспечения#cite_note-1
12. Никлаус Вирт Долой "жирные" программы Открытые системы №06, 1996. <http://www.osp.ru/os/1996/06/179017/>
13. Антивирусный прогноз погоды: облачно <http://www.infosecurity.ru/cgi-bin/mart/arts.pl?a=101001> — 2011
- 14.Полноценные хранилища свободных знаний - Википедия и Проект Гутенберг <http://rus-linux.net/lib.php?name=/MyLDP/freesoft/impossible-2.html>
- 15."Лаборатория Касперского": экскурсия и взгляд изнутри http://www.thg.ru/business/kaspersky_tour/print.html
- 16.Антивирусные вендоры ищут выход из технологического тупика http://www.anti-malware.ru/antivirus_trends
17. Семенов Ю.А. (ГНЦ ИТЭФ) Направления и тенденции развития ИТ-технологий <http://book.itep.ru/4/7/resources.htm>
18. Количество строк кода http://ru.wikipedia.org/wiki/Количество_строк_кода
19. Почти объективно на тему «чем FreeBSD лучше Linux» <http://eax.me/freebsd-vs-linux/>

Материалы поступили 19.03.2012, опубликовано в Интернет 20.04.2012 по положительной рецензии д.т.н. Иванова А.И. (Пенза)