

# ОСОБЛИВОСТІ АВТОМАТИЗАЦІЇ ПРОГНОЗНОГО АНАЛІЗУ ДАНИХ ПІД УПРАВЛІННЯМ ОПЕРАЦІЙНИХ СИСТЕМ LINUX



В. І. Дубук, канд. техн. наук,  
Т. З. Цюра

Досягнення сучасних інформаційних систем і технологій (ІСТ) широко використовуються на практиці під час моделювання різних економічних явищ і процесів. Особливо важливе значення як з точки зору теорії, так і практики математичного моделювання і комп'ютерного симулювання, економетричного аналізу для прийняття високоефективних управлінських рішень відіграє розв'язання задач прогнозування.

При цьому можуть використовуватися різні комп'ютерно-орієнтовані методи, частина яких ґрунтується на застосуванні програмного забезпечення інструментальних засобів аналізу даних [2; 6]. Інша група методів побудована на основі застосування програмного забезпечення табличних процесорів [1; 3; 4], серед яких найширше використовуються програми Microsoft Excel, Star Office Calc, Open Office.org Calc, Gnumeric, KSpreadsheet, ABS та ін. [1; 3–5; 7]. Деякі з перелічених програмних засобів розроблені і використовуються в інформаційних системах (ІС) виключно під управлінням операційних систем (ОС) сімейства Microsoft Windows [1], інші – являють собою кросплатформені версії програмного забезпечення [7], які успішно працюють під управлінням різнотипних операційних систем – як із сімейства Microsoft Windows, так і Unix/Linux.

Для розв'язання задач прогнозного аналізу даних на основі наявних множин статистичних числових даних можна застосувати засоби табличного процесора Gnumeric [7]. Цей програмний засіб працює в ІС під управлінням ОС з сімейства Linux та Microsoft Windows і має вбудовану групу потужних інструментів для статистичного аналізу даних: розраховує ковзне середнє значення, здійснює експоненціальне згладжування, обчислення регресій, містить значну кількість функцій різного призначення [10–12]. Крім того, програмний засіб Gnumeric володіє вільною ліцензією на використання, тому його інсталяція і практичне використання є економічно ефективним.

Про практичне застосування методу прогнозування на основі табличного процесора Microsoft Excel ідеться в роботі [1]. Однак розв'язання проблеми

прогнозного аналізу даних у цій роботі остаточно не вирішене в частині розробки ефективних методів аналізу даних засобами ІС, що працюють під управлінням операційних систем сімейства Linux. Розв'язання вказаної частини проблеми є актуальним і становить практичну і теоретичну цінність.

У даній науковій праці розробляються методи розв'язання задач прогнозного аналізу економіко-статистичних даних на основі прикладного програмного забезпечення Gnumeric, що працює в ІС під управлінням ОС сімейства Linux.

Аналіз здійснюється засобами застосування вбудованої функції обчислення середнього значення AVERAGE(), розрахунку ковзного середнього значення засобами відповідних інструментів «ковзне середнє» та «експоненціальне згладжування» табличного процесора Gnumeric [7].

На основі застосування вказаних засобів ґрунтуються методи прогнозного аналізу даних з визначенням ковзного середнього значення. Вони уможливають спрогнозувати значення тренду часового ряду функціональної залежності на один часовий відлік уперед.

**Метод розрахунку** на основі визначення ковзного середнього значення застосовується на практиці для згладжування і прогнозування майбутніх значень функціональної залежності на основі числових даних попередніх значень.

Часовий ряд числових даних є множиною пар даних  $(t, y)$ , в якій  $t$  – моменти, значення часу (незалежні змінні), а  $y$  – параметр, що характеризує значення функції досліджуваного процесу (функціонально часово-залежна змінна).

Практичне застосування методу визначення ковзного середнього значення уможливило виявити тенденції зміни фактичних значень параметра  $y$  в часі і визначити прогнозні значення  $y$  в майбутні часові періоди.

Побудовану прогнозу математичну модель можна ефективно застосовувати у випадках, коли для значень прогнозованого параметра  $y$  спостерігається стійка тенденція в динаміці процесу часової залежності.

Водночас даний метод є слабоефективним у випадках порушення стійкої тенденції динаміки процесу. Наприклад, для моделювання соціально-економічних процесів чи явищ під час військових дій, стихійних лих, суспільних революцій та заворушень, а також у разі різкої, значної зміни внутрішніх параметрів соціально-економічної системи або параметрів зовнішнього оточуючого середовища (рівень інфляції, ціни на сировину, законодавчі, нормативні та регулюючі акти і т. п.); миттєвої зміни управлінських дій підприємства тощо.

Основа методу полягає в заміні фактичних значень досліджуваного часового ряду даних середніми значеннями, які уникають випадкових коливань числових даних. При цьому отримують згладжений ряд значень досліджуваного параметра, який уможливило визначити головну тенденцію його часової зміни.

Метод ґрунтується на зображенні прогнозного значення  $y_i^*$  у вигляді середнього значення  $m$  попередніх спостережних значень  $y_{t=i}(i = \overline{1, m})$ , або:

$$y_i^* = \frac{1}{m} \sum_{i=1}^m y_{t=i} \quad (1)$$

Якщо, наприклад, у разі дослідження часового ряду даних про витрати підприємства за місяцями як прогнозне значення вибрати ковзне середнє за два попередні місяці ( $m = 2$ ), то прогнозом на третій місяць (березень) буде середнє значення показників за два попередні місяці (січень, лютий). Якщо ж розрахувати шестимісячне ковзне середнє значення ( $m = 6$ ), то прогнозом на сьомий місяць (липень) буде середнє значення показників за шість попередніх місяців (січень, лютий, березень, квітень, травень, червень).

У літературі відсутні критерії для визначення оптимальної кількості  $m$  членів для розрахунку ковзного середнього, проте простежується ефект: чим менше  $m$ , тим сильніше прогнозна оцінка даних реагує на коливання часового ряду, а для більших значень  $m$

процес прогнозування стає більш інерційним.

У практичних розрахунках значення  $m$  приймають з рівності:

$$m = 2 \dots 10 \quad (2)$$

За наявності достатньої кількості елементів часового ряду значення  $m$  можна визначити шляхом застосування відповідного покрокового алгоритму.

*Крок 1.* Задати декілька попередніх значень  $m$ .

*Крок 2.* Згладити часовий ряд, використовуючи кожне задане значення  $m$ .

*Крок 3.* Обчислити середню відносну похибку прогнозування за формулою:

$$e = \frac{1}{m} \sum_{i=1}^k \left| \frac{y_t - y_t^*}{y_t} \right| \quad (3)$$

де  $k$  – кількість відліків моментів часу  $t_i$ .

*Крок 4.* Вибрати значення  $m$ , яке відповідає мінімальній відносній похибці прогнозування.

Реалізацію процесу згладжування і прогнозування за методом ковзного середнього значення в середовищі програмного засобу Gnumeric можна здійснити різними засобами:

- введенням у комірки електронної таблиці відповідної формули, наприклад, з використанням вбудованої функції AVERAGE();

- за допомогою інструмента «ковзне середнє» значення опції «інструменти – статистичний аналіз – прогнозування»;

- за допомогою інструмента «експоненціальне згладжування» опції «інструменти – статистичний аналіз – прогнозування».

*Приклад 1.* Враховуючи дані таблиці щомісячних витрат підприємства за 11 місяців поточного року, обчислити прогнозне значення витрат за грудень (дванадцятий) місяць року.

**Алгоритм розв'язання прикладу**

Таблиця 1

Місяць	1	2	3	4	5	6	7	8	9	10	11	12
Витрати, тис. грн.	6500,00	7100,00	7350,00	7100,00	6800,00	6500,00	6600,00	7200,00	7800,00	8250,00	8400,00	X

Процес одержання згладженого часового ряду, а також прогноз про витрати підприємства на грудень місяць поточного року за даними вихідного часового ряду для попередніх одинадцяти місяців може здійснюватися за таким алгоритмом.

*Крок 1.* На основі даних (дивись табл.1) на робочому листі Gnumeric створюється таблиця, яка заповнюється числовими даними вихідного часового ряду.

*Крок 2.* Формуються і заносяться в таблицю числові дані згладжених часових рядів для двох-, трьох- та чотирьохмісячного ковзного середнього значення.

*Крок 3.* Будуються графіки вихідного часового ряду і згладжених рядів.

*Крок 4.* За виразом (3) обчислюються середні відносні відхилення отриманих згладжених часових рядів від вихідного часового ряду.

*Крок 5.* Як модельний вибирається згладжений часовий ряд з мінімальним середнім відносним відхиленням, і на основі його показників складається прогноз про витрати підприємства на дванадцятий місяць поточного року.

**Метод розв'язання прикладу**

1. Заповнюємо діапазон клітинок (A5:B15) робочого листа A1 даними часового ряду з табл. 1.

2. За даними часового ряду з діапазону клітинок (A5:B15) будуємо на основі методу розрахунку ковзного середнього значення три табличні моделі досліджуваної залежності за даними для двох, трьох і чотирьох місяців відповідно. Значення отриманих згладжених часових рядів розміщуємо відповідно в діапазоні клітинок: (C7:C16), (D8:D16), (E9:E16).

2.1. Спочатку будуємо ряд значень ковзного середнього за двома місяцями: в клітинку C7 вводимо формулу  $=AVERAGE(B5:B6)$  і маркером заповнення копіюємо її в діапазон комірок (C8:C16). При цьому діапазон (C7:C16) заповнюється обчисленими показниками для двомісячного середнього значення.

Подібним чином будуються ряди значень для трьох- та чотирьохмісячного ковзного середнього значення.

2.2. У клітинку D8 вводимо формулу  $=AVERAGE(B5:B7)$  і маркером заповнення копіюємо її в діапазон клітинок (D9:D16). При цьому діапазон (D8:D16) заповнюється показниками трьохмісячного ковзного середнього значення.

2.3. У клітинку E9 вводимо формулу  $=AVERAGE(B5:B8)$  і маркером заповнення копіюємо її в діапазон клітинок (E10:E16). При цьому діапазон клітинок (E9:E16) заповнюється показниками чотирьохмісячного ковзного середнього значення.

На рис. 1 наведено графік вихідного часового ряду і побудовані відносно нього прогностичні лінії трендів ковзного середнього значення.

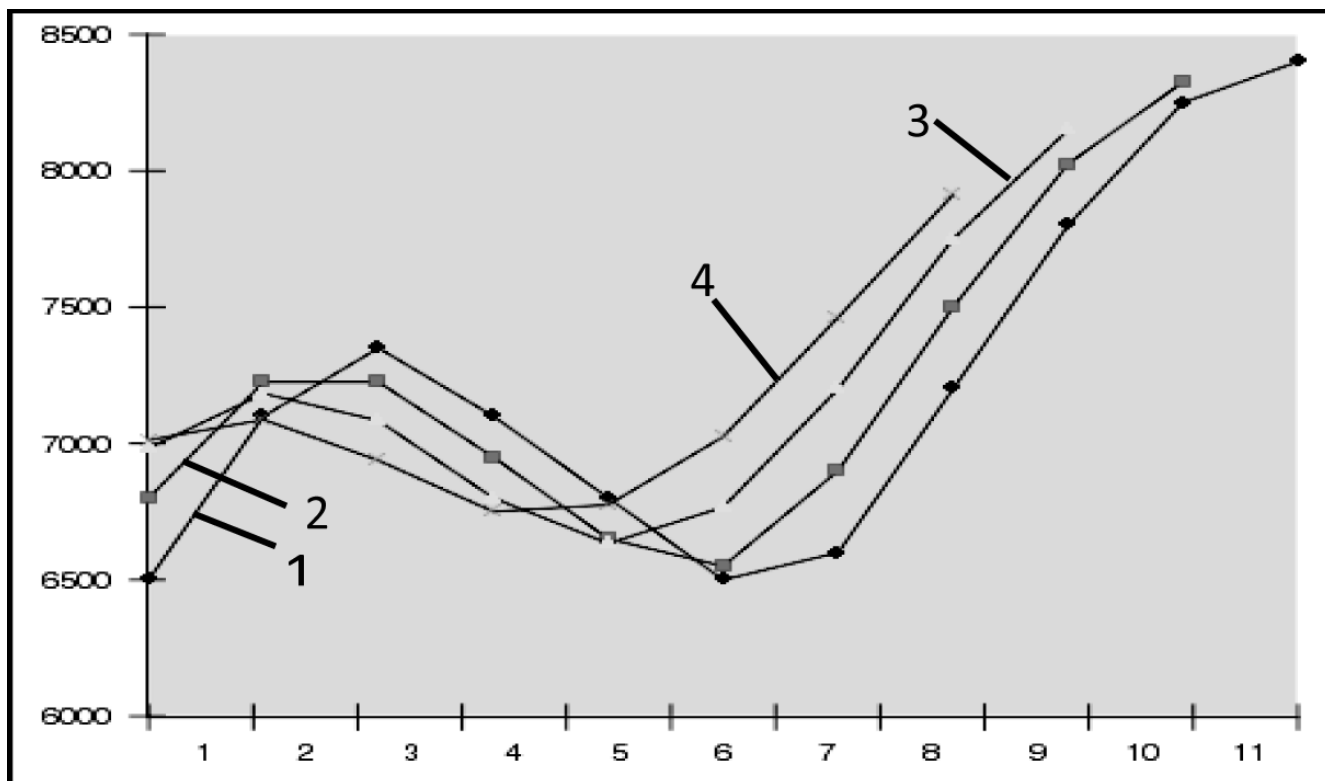


Рис. 1. Графічні залежності вихідного часового ряду (1) і згладжених часових рядів.  
Динаміка витрат підприємства протягом року:  
2 – за двома місяцями; 3 – за трьома місяцями; 4 – за чотирма місяцями

Оскільки отримані значення згладжених часових рядів на основі ковзного середнього значення базуються на даних, отриманих з попередніх спостережень, то вони запізнюються порівняно з відповідними значеннями вихідного часового ряду: лінії тренду ковзного середнього значення зсунуті відносно графіка вихідного часового ряду.

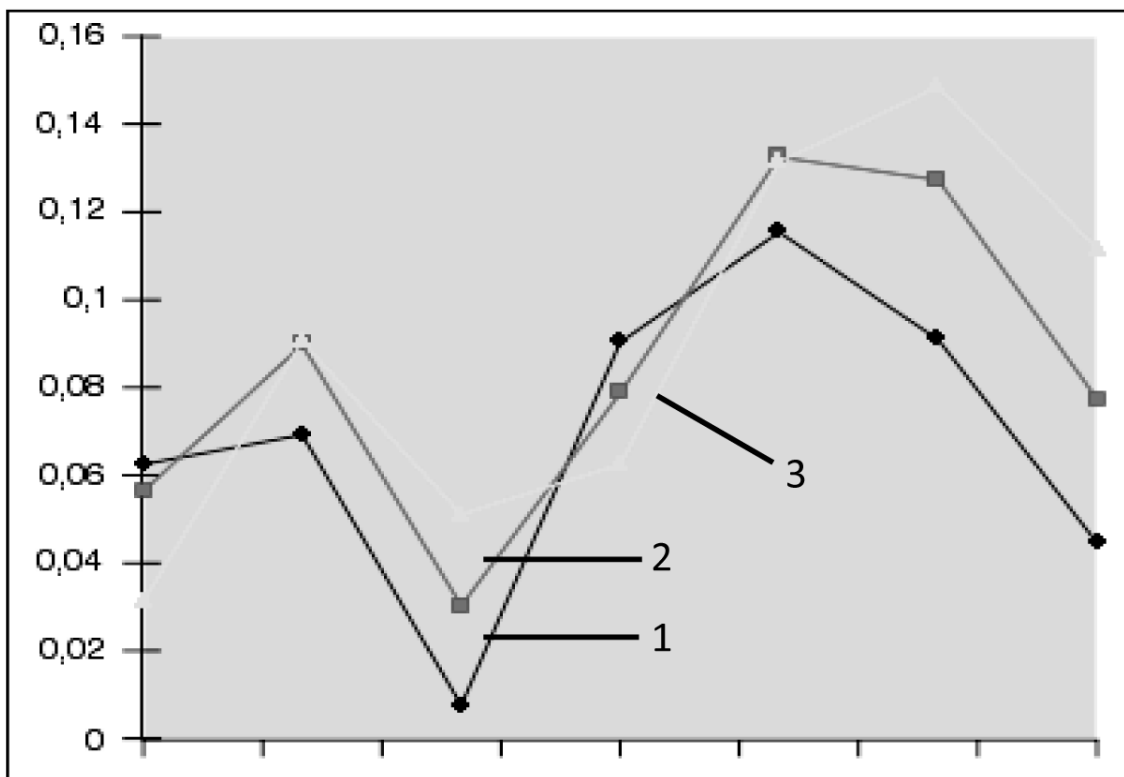
У табл. 2 наведені відносні відхилення значень для двох-, трьох- та чотирьохмісячного ковзного середнього значення від відповідних значень вихідного

часового ряду.

Для виконання порівняльного аналізу похибок для двох-, трьох-, чотирьохмісячного ковзного середнього значення було взято однакову кількість спостережень (рис. 2; 3).

### Висновки

З аналізу наведених табличних даних і графічних залежностей відносних відхилень можна стверджувати, що для згладжування вихідного часового ряду і формулювання прогнозу про тенденції зміни витрат



**Рис.2. Відносні відхилення розрахункових значень, розраховані: 1 – за двома місяцями; 2 – за трьома місяцями; 3 – за чотирма місяцями**

підприємства найприйнятнішою є модель на основі двомісячного ковзного середнього значення. Така модель найточніше реагує на коливання вихідного часового ряду, має менші відносні похибки прогнозування і уможливує визначити прогнозне значення витрат підприємства на черговий наступний місяць.

Порівняння середніх відносних похибок надає підстави вважати модель двомісячного ковзного се-

реднього значення адекватнішою для згладжування і прогнозування, оскільки її застосування в усіх точках розглянутого часового діапазону дає менші похибки.

У роботі розроблені методи прогнозного аналізу економіко-статистичних даних, які ґрунтуються на застосуванні вбудованої функції обчислення середнього значення AVERAGE(), розрахунку ковзного середнього значення засобами відповідного інструменту «ковзне

*Таблиця 2*

*Таблиця значень відносних відхилень*

Місяць	Відносні відхилення, %		
	за двома місяцями	за трьома місяцями	за чотирма місяцями
1	2	3	4
3	7,48	н/д	н/д
4	1,76	1,64	н/д
5	6,25	5,64	3,13
6	6,92	8,97	9,04
7	0,76	3,03	5,11
8	9,03	7,87	6,25
9	11,54	13,25	13,14
10	9,09	12,73	14,85
11	4,46	7,74	11,16
Середнє відносне відхилення	6,37	7,61	8,95

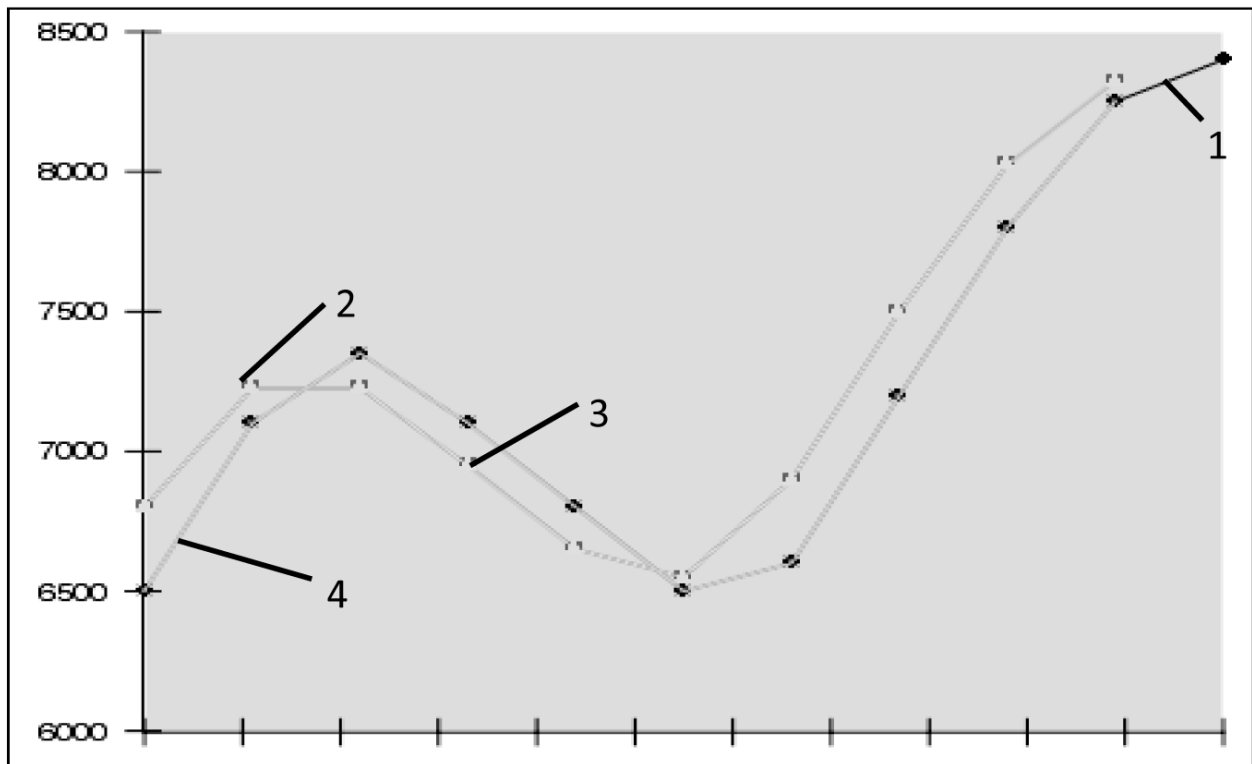


Рис.3. Графіки вихідної функціональної залежності (1); ковзного середнього значення за двома місяцями, розрахованого з використанням функції AVERAGE() (2), інструмента «ковзне середнє»(3), інструмента «експоненціальне згладжене» (4).

Динаміка витрат підприємства за три квартали біжучого року:

- 1 – фактичні витрати підприємства за кожний місяць, тис. грн.
- 2 – розраховані значення за визначенням ковзного середнього значення за двома місяцями (розраховані за функцією AVERAGE()), тис. грн.;
- 3 – розраховані значення за визначенням ковзного середнього значення за двома місяцями, (отримані з допомогою інструменту статистичного аналізу «ковзне середнє»), тис. грн.;
- 4 – розраховані значення за визначенням ковзного середнього значення за двома місяцями, (отримані з допомогою інструменту «експоненціальне згладжене»), тис. грн.

середнє»), та інструменту «експоненціальне згладжене»), табличного процесора Gnumeric.

Розроблені методи уможливають прогноз значень тренду часового ряду функціональної залежності на один часовий відлік вперед. Існує можливість розробки методів прогнозування більш віддалених, наступних у часі значень функції.

Розроблені в роботі методи розв'язання задач прогнозного аналізу економіко-статистичних даних на основі прикладного програмного забезпечення Gnumeric під управлінням ОС сімейства Linux є ефективними і рекомендуються для практичного використання.

#### ЛІТЕРАТУРА

1. Гавриленко В. В., Пархоненко А. М. Решение задач аппроксимации средствами Excel // Компьютеры + программы, 2002. – №12. – С. 42–47.  
 2. Дубук В. І. Використання інструментальних засобів аналізу даних на персональних комп'ютерах під управлінням операційної системи ASP Linux // Інформац. технології в сучасній

економіці, менеджменті та освіті: Матер. міжнар. наук.-пр. конф., Львів, 11-12.02.2005 р. – Львів: Львів. філія Європ. ун-ту, 2005. – С. 57–59.

3. Дубук В. І. Автоматизований аналіз даних на персональних комп'ютерах // Науково-технічна інформація. – 2008. – № 3(37). – С. 44–45.

4. Дубук В. І., Цюра Т. З. Автоматизація аналізу даних засобами інформаційних систем під управлінням операційних систем Linux // Тези доп. Міжвуз.наук.-пр. конф. «Сучасні інформаційні технології в економіці, менеджменті та освіті», Львів, 11.12.2009 р. – Львів: Львів. філія Європейського університету, 2009.

5. Тюрин Ю. Н., Макаров А. А. Анализ данных на компьютере / Под ред. В. Э. Фигурнова. – М.: ИНФРА-М, 2003. – 544 с.

6. Hornik K. The R FAQ // <http://www.ci.tuwien.ac.at/~hornik/R/>  
 7. Baudais E., Breit K., Custer A., Canty T., Dassen R., Goldberg J., Guelzow A.J., Hellan J.K., De Icaza M., Livonen J-P., Kirillov A., Klost S., Leblanc G., Luangker-son L., Miesbauer T., Schuller W., Tigelaar A.S., Twardy Ch., Weber A., Welinder M. The Gnumeric Manual. Ver.1.9.16 – The Gnumeric Team, Gnome Documentation Project, 1998-2009 // <http://www.gnumeric.org>