

# Comparative Performance of XSLT Processors in Cocoon

John Berry  
Vadim Gritsenko  
Jack Ivers  
Franqueli Mendez  
Scott Roth

# The testbed: Merck Manual Second Home Edition

---

- <http://www.merck.com/mmhe/index.html>
- <http://mmh.banyu.co.jp/mmhe2j/index.html>
- Free, on-line version of the widely-recognized book
  - Content complete
  - Includes additional multimedia features
- Two of Merck's highest traffic sites
- Single source XML publishing: same XML content drives production of printed book and website
  - Adobe FrameMaker used to maintain the XML master

THE  
**MERCK MANUAL**  
OF  
**MEDICAL INFORMATION**  
SECOND HOME EDITION ONLINE

|         |  |
|---------|--|
| SECTION | <a href="#">Lung and Airway Disorders</a>  |
| CHAPTER | <a href="#">Infiltrative Lung Diseases</a>   |
| TOPICS  | <a href="#">Introduction</a> · <a href="#">Cryptogenic Organizing Pneumonitis</a> · <a href="#">Desquamative Interstitial Pneumonia</a> · <a href="#">Idiopathic Pulmonary Fibrosis</a> · <a href="#">Langerhans' Cell Granulomatosis</a> · <a href="#">Lymphoid Interstitial Pneumonia</a> · <a href="#">Pulmonary Alveolar Proteinosis</a> · <a href="#">Sarcoidosis</a> |

Merck & The Merck Manuals  
Committed to Providing  
Medical Information »

Search *The Second Home Edition,*  
Online Version

Index

|   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|
| A | B | C | D | E | F | G | H | I |
| J | K | L | M | N | O | P | Q | R |
| S | T | U | V | W | X | Y | Z |   |

Sections

- [Accidents and Injuries](#)
- [Blood Disorders](#)
- [Bone, Joint, and Muscle Disorders](#)
- [Brain, Spinal Cord, and Nerve Disorders](#)
- [Cancer](#)

## Sarcoidosis

*Sarcoidosis is a disease in which abnormal collections of inflammatory cells (granulomas) form in many organs of the body.*

The cause of sarcoidosis is unknown. It may result from an infection or from an abnormal response of the immune system. Inherited factors may be important. Sarcoidosis develops predominantly between the ages of 20 and 40 and is most common among Swedes and American blacks, although it can occur in anyone.

Sarcoidosis is characterized by the presence of collections of inflammatory cells (granulomas). The disease is primarily one of the lungs, but granulomas also form in the lymph nodes, lungs, liver, eyes, and skin, and less often in the spleen, bones, joints, skeletal muscles, kidneys, heart, and nervous system. The granulomas may eventually disappear completely or become scar tissue.

### Symptoms


Many people with sarcoidosis have no symptoms, and the disease is discovered during a chest x-ray that is taken for other reasons. Most people develop minor symptoms that do not progress. Serious symptoms are rare.

The symptoms of sarcoidosis vary greatly, according to the site and

Buy The Book



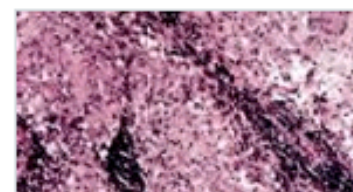
 Print This Topic


 Email This Topic

### Pronunciations

-  [angina](#)
-  [angiotensin](#)
-  [bronchi](#)
-  [bronchoscopy](#)
-  [computed tomography](#)

### Multimedia



 Lymph Node Specimen: Sarcoidosis

万有製薬 | メルクマニュアル

最新  
メルクマニュアル  
医学百科  
家庭版

米国メルク社とメルクマニュアル

メルクマニュアルとは...  
医学情報提供の歴史 >>

 メルクマニュアル家庭版  
を検索

検索

索引

|   |   |   |   |   |   |    |   |   |   |
|---|---|---|---|---|---|----|---|---|---|
| あ | い | う | え | お | か | き  | く | け | こ |
| さ | し | す | せ | そ | た | ち  | つ | て | と |
| な | に | ぬ | ね | の | は | ひ  | ふ | へ | ほ |
| ま | み | む | め | も | や | ゆ  | よ |   |   |
| ら | り | る | れ | ろ | わ | 記号 |   |   |   |

セクション

知っておきたい基礎知識

薬についての基礎知識

心臓と血管の病気

肺と気道の病気

|       |   |
|-------|---|
| セクション | 肺と気道の病気   |
| 章     | 浸潤性肺疾患  |
| トピック  | はじめに・特発性肺線維症・剥離性間質性肺炎・リンパ性間質性肺炎・特発性器質化肺炎・ランゲルハンス細胞肉芽腫症・サルコイドーシス・肺胞タンパク症 |

## サルコイドーシス

サルコイドーシスとは、体のさまざまな器官において炎症細胞が異常に蓄積し、肉芽腫を生じる病気です。

サルコイドーシスの発症原因はわかっていません。感染症もしくは免疫系の異常反応によって発症するのではないかと考えられています。遺伝的な要素が重要という可能性もあります。サルコイドーシスは20～40歳の間に発症することが多く、人種を問わず発症しますが、最も多くみられるのはスウェーデン人とアフリカ系アメリカ人です。

サルコイドーシスは、肉芽腫という炎症細胞の蓄積がみられることが特徴です。この病気は最初、片方の肺に生じますが、肉芽腫はリンパ節、両方の肺、肝臓、眼、皮膚にも形成されます。まれに心臓、骨、関節、骨格筋、腎臓、心臓、神経系にも形成されます。肉芽腫はやがて完全になくなるが、瘢痕化した組織になります。

### 症状

サルコイドーシスの多くは何も症状がないため、他の理由で実施した胸部X線検査によって発見されます。ほとんどの患者に軽い症状がみられますが、これは進行しません。重い症状が現れるのはまれです。

サルコイドーシスの症状は、病変のある部位や範囲によって大きく異なります。発熱、疲労感、鈍い胸痛、全身のけん怠感、体重減少、関節痛などは、患者の約3分の1で最初にみられる徴候です。リンパ節の腫大は一般的ですが、何も症状は起こしません。発熱や寝汗は、病気が治るまで繰り返すこともあります。

サルコイドーシスによって最も影響を受ける器官は肺です。肺と心臓が接する部分には、右側の気管支のリンパ節の腫大が、胸部X線検査で目撃することが

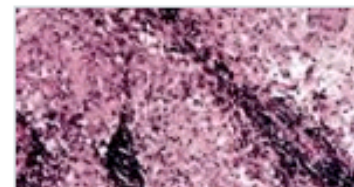
書籍について



印刷用画面

メールで紹介

マルチメディア



リンパ節標本：サルコイドーシス

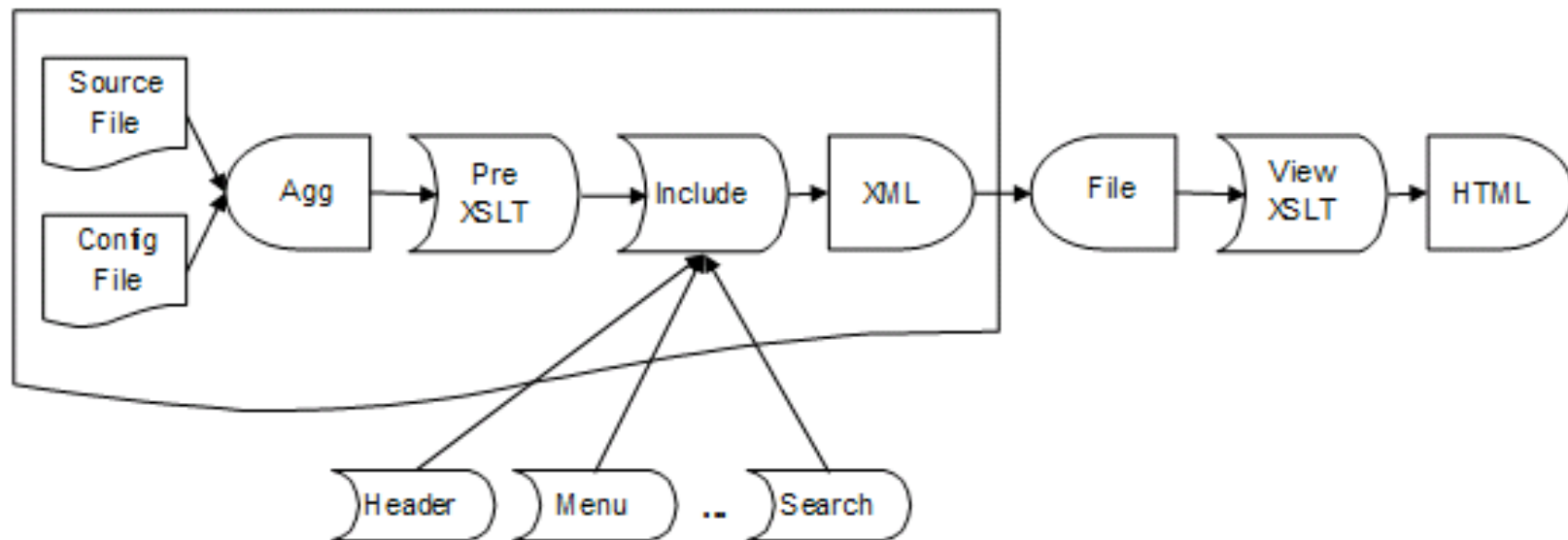
# Cocoon site wins 2005 Webby Award!

<http://www.webbyawards.com/webbys/current.php>

| <b>HEALTH</b>  |                                 |
|--|---------------------------------|
| <b>WEBBY AWARD WINNER</b>  | <b>CREDITED ORGANIZATION</b>    |
| † The Merck Manual -- Second Home Edition Online<br><a href="http://www.merckhomeedition.com">http://www.merckhomeedition.com</a>  | <i>Merck &amp; co, Inc</i>      |
| <b>PEOPLE'S VOICE WINNER</b>   |                                 |
| † HIV STOPS WITH ME<br><a href="http://www.hivstopswithme.org">http://www.hivstopswithme.org</a>                                   | <i>Better World Advertising</i> |
| <b>NOMINEES</b>  |                                 |
| † AOL Diet & Fitness, AOL Keyword: Diet<br><a href="http://aolsvc.aol.com/dietfitness...">http://aolsvc.aol.com/dietfitness...</a> | <i>America Online, Inc.</i>     |
| † Epilepsy.com<br><a href="http://www.epilepsy.com">http://www.epilepsy.com</a>  | <i>The Epilepsy Project</i>     |
| † Heart of New Life<br><a href="http://www.heartofnewlife.com">http://www.heartofnewlife.com</a>                                   | <i>Brilliant Blue LLC</i>       |
| † HIV STOPS WITH ME<br><a href="http://www.hivstopswithme.org">http://www.hivstopswithme.org</a>                                   | <i>Better World Advertising</i> |
| † The Merck Manual -- Second Home Edition Online<br><a href="http://www.merckhomeedition.com">http://www.merckhomeedition.com</a>  | <i>Merck &amp; co, Inc</i>      |

# Application details

- All pages are dynamically rendered
- 20MB XML transformed into 4000+ topic-level XML chunks, 10-100K in size, as part of an off-line process (Cocoon CLI)
  - Plus a few larger chunks: index, etc.
- Rather complex aggregate page-rendering pipeline:



# Trouble

---

- MMHE2 application easily passed rigorous lab tests
- Application was then launched in production; but began to fail with Java OOM errors
  - Irregularly, every 3-10 days
  - Soft correlation with load
- Unable to reproduce failures in any test environment
- Crisis!

# Diagnostics

---

- **Multiple rounds of optimization / tweaking / re-launch**
  - Greatly improved pipeline cacheability
  - Reduced size of large XML files, eliminated some large index pages
  - But still OOM ...
- **Enabled core logging on production so we could observe memory & store janitor behavior**
  - Observed “cache crashes” and “memory gobbles”
- **XSLT style sheet validation**
  - Worked with Jacek Ambroziak of Ambrosoft, developer of Gregor, to see if our XSLTs were inherently flawed
  - Found and corrected a few issues, but nothing large and no impact on OOMs

# Breakthrough

---

- Finally reproduced failure on test server
  - After weeks of trying, we were able to reproduce OOM error on test server by applying extremely heavy loads
- In-depth memory profiling of all code, not just ours, uncovered serious memory problems in Xalan 2.4
  - Heavy memory utilization in multi threaded environment due to Xalan not always releasing its internal object pools
- Because of shared configuration restrictions, we were unable to implement Xalan 2.6
- The only option was to explore other XSLT processors

# Initial testing: February 2005

---

- We tested Xalan 2.4\* and Saxon 7 (JDK 1.4)
- Because problem only visible under load, testing was conducted under load
  - Webstress tool used to generate simulated multi-user load
  - 1000s of unique URLs taken from production clickstream
  - Pipeline caching disabled to force heavy XSLT processing
- Primary measure was “garbage collection velocity” or “memory churn”
  - Enable verbose GC
  - Watch the catalina output for [GC ...] and [Full GC ...] messages
  - Watch frequency of messages and % of full GCs
- Unscientific in some ways but seemed to bear most directly on our “OOM under load” problem

\*As bundled with JDK 1.4

# Initial results

---

- Xalan 2.4 memory churn was extremely high
- Saxon 7 was far better
  - Much lower memory churn
  - Did require XSLT changes: less forgiving of XSLT sloppiness, some namespace related quirks
- We shifted to Saxon 7 in production, OOM errors went away

# Sept 2005 testing

---

- **Goals**
  - Better, more scientific tests, conducted under less time pressure
  - Good enough to publish and share with the Cocoon community
- **XSLT processor / JVM configurations tested**
  - On JDK 1.4 (primarily to confirm earlier test results)
    - Xalan 2.4\*
  - On JDK 1.5
    - Xalan 2.7
    - Saxon 7
    - Saxon 8
    - XSLTC
  - Xalan serializer used in all cases
- **We plan to test Gregor and publish the results after GT**

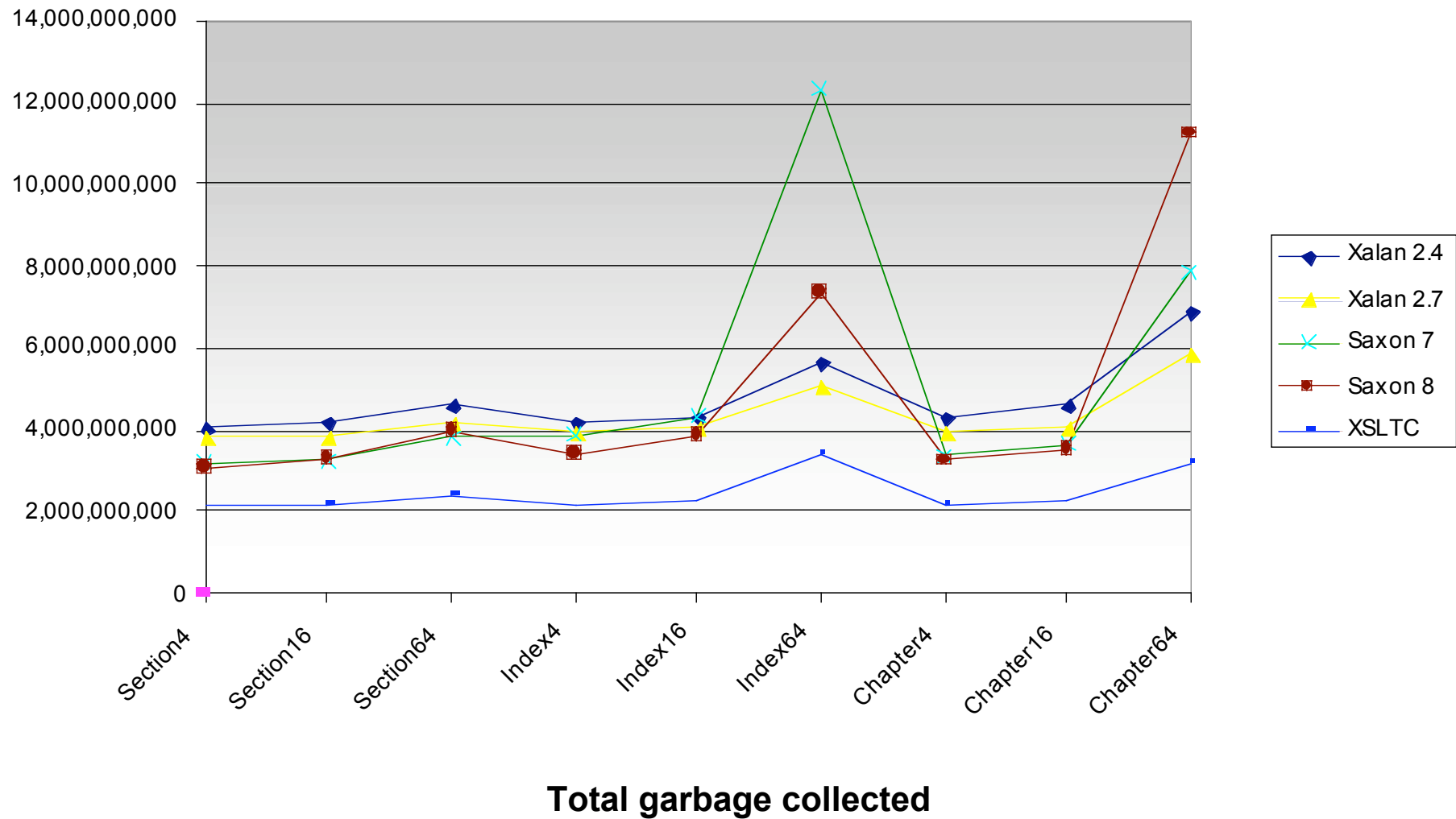
\*As bundled with JDK 1.4

# Sept 2005 test methodology

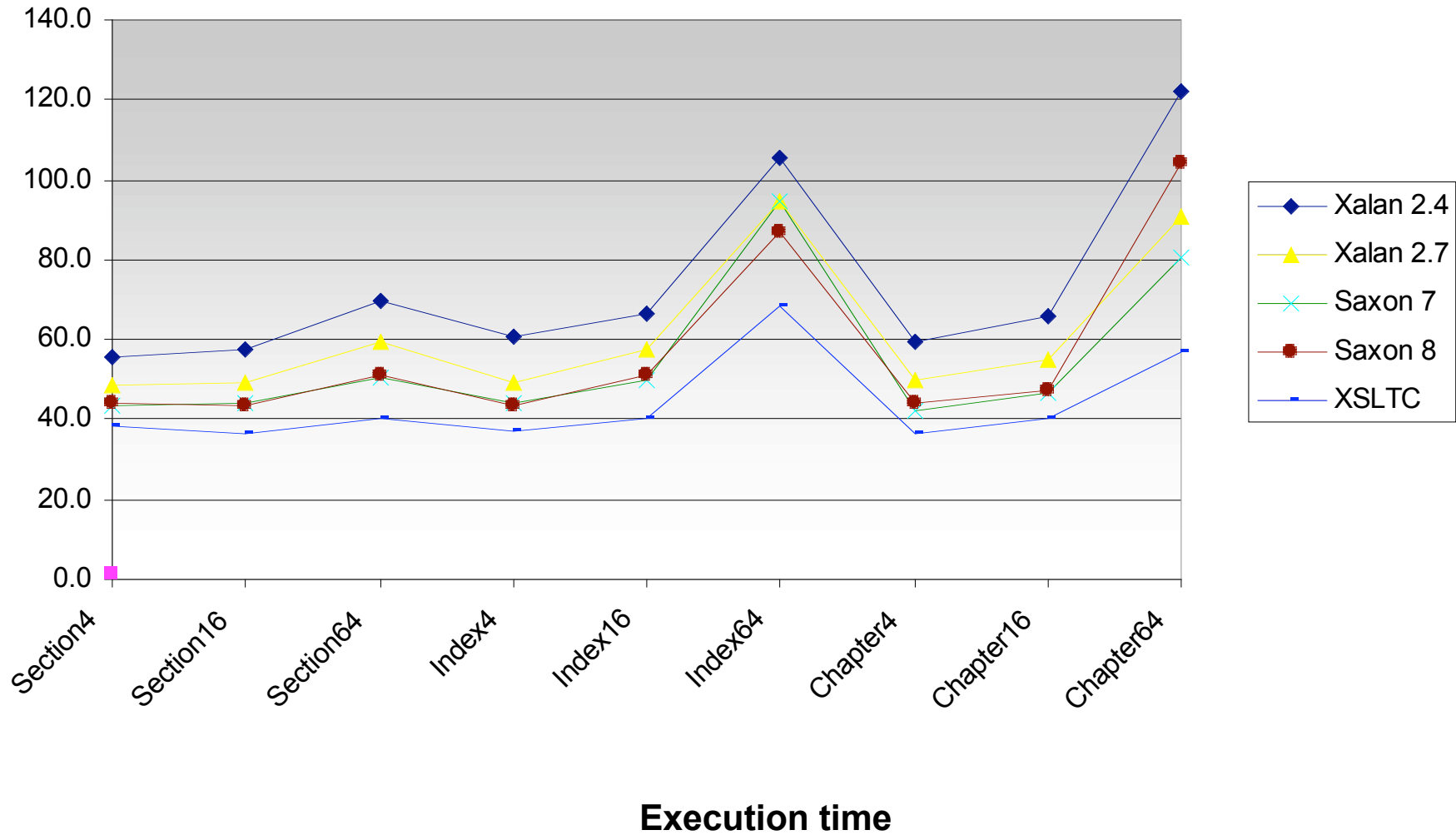
---

- More scientific multi-user “memory churn” test
  - Placed a ~4 pageview/second load on the application (single server, 256M JVM, pipeline caching disabled)
  - Measurements:
    - Total garbage collected
    - CPU utilization
    - Average response time
    - Memory footprint
- Added a single user test
  - Using Apache Bench, we ran single-user 1000 iteration tests on specific URLs
  - Measured:
    - Elapsed time (processing time)
    - Total garbage collected
    - Memory footprint

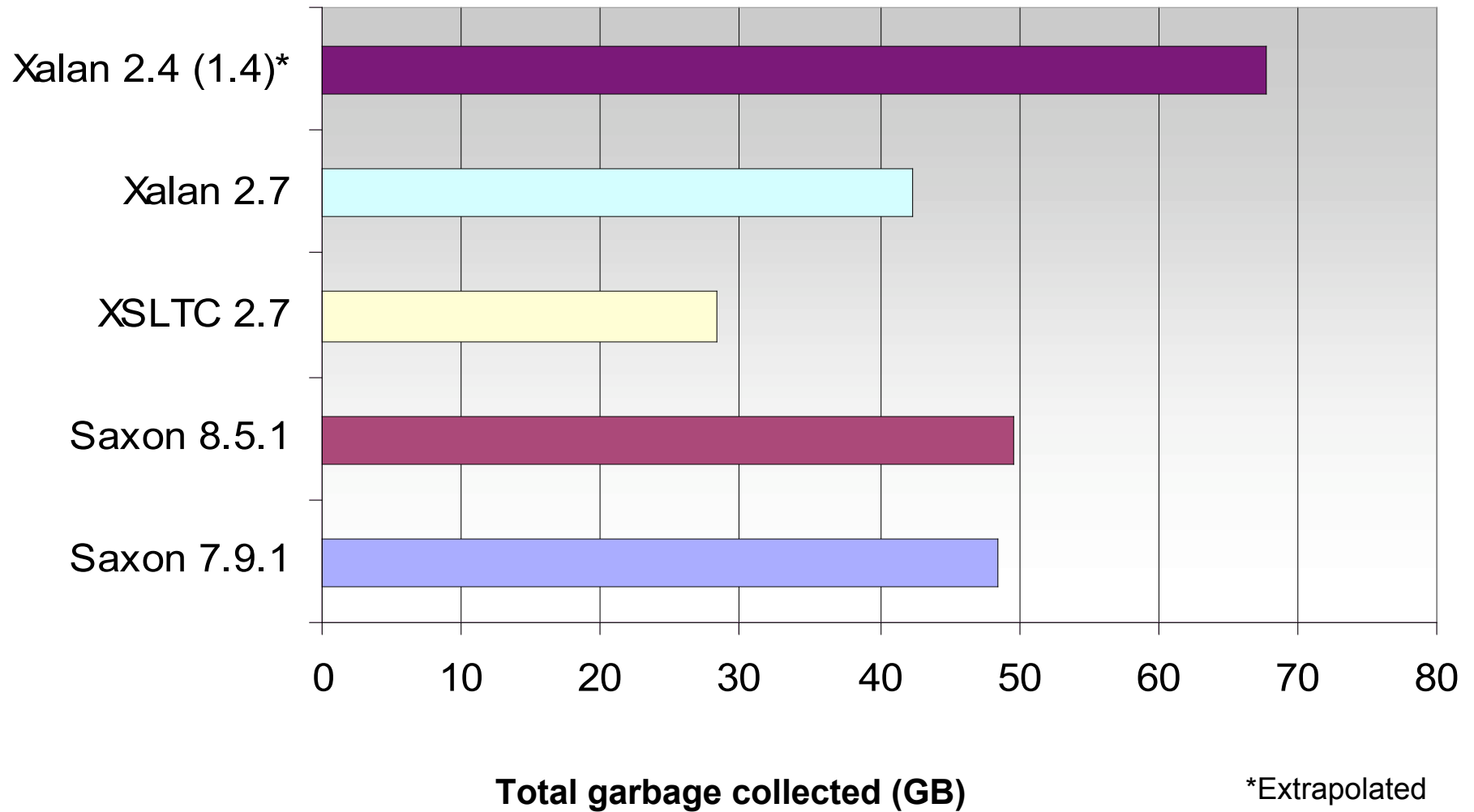
# Results: Single-user: garbage collected



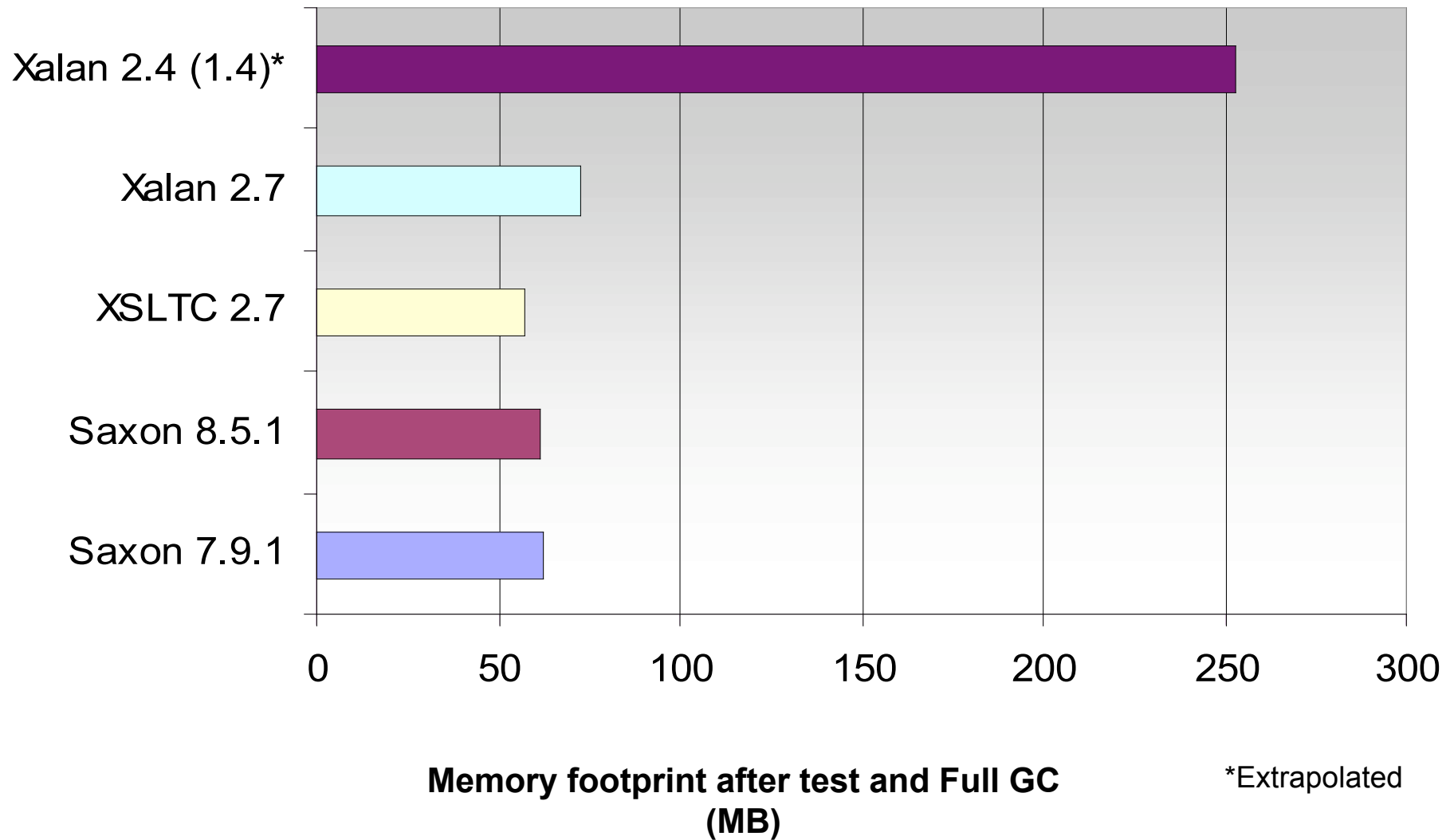
# Results: Single-user: execution time



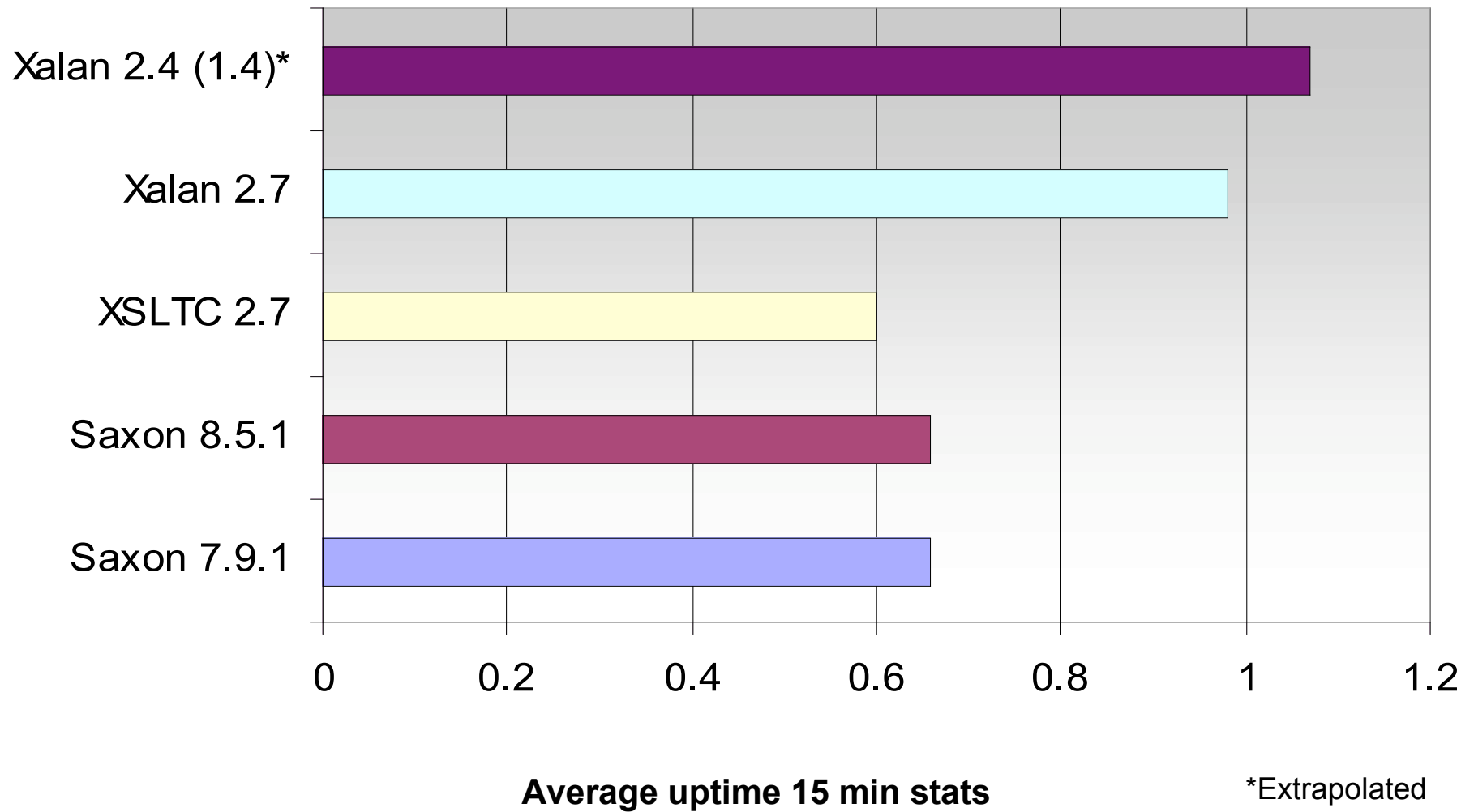
# Results: Multi-user: garbage collected



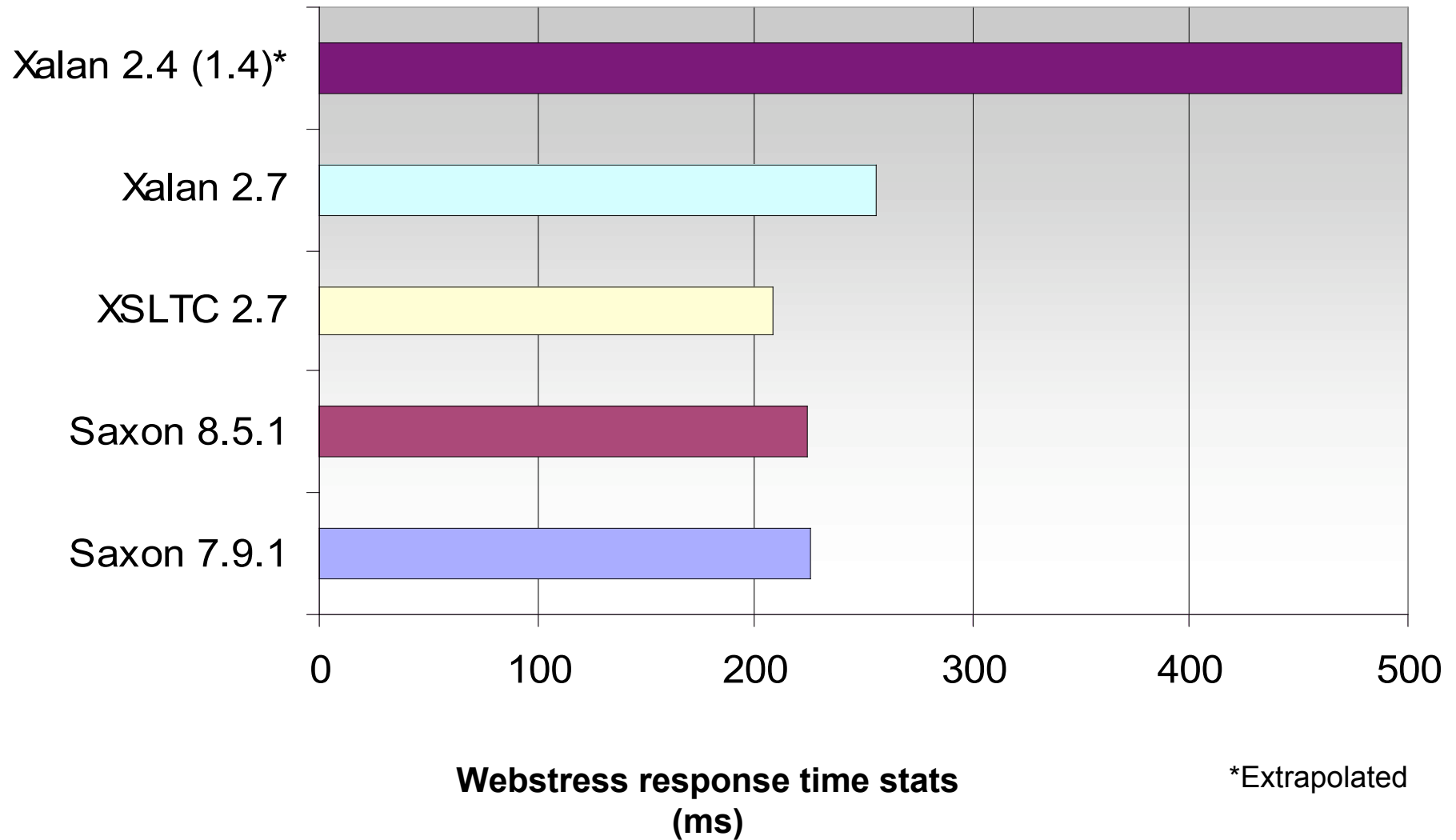
# Results: Multi-user: memory footprint



# Results: Multi-user: CPU utilization



# Results: Multi-user: response time



# XSLT processors compared

---

- **Xalan**
  - 2.4 was a seriously flawed release
  - 2.7 appears to have corrected many of these deficiencies
    - No longer lags Saxon significantly
  - Xalan serializer is superior to Saxon's: faster, with desired behavior on disable-output-escaping
- **Saxon**
  - XSLT changes required due to more strict syntax enforcement
  - Both V7 and V8 appear to have case-specific inefficiencies in some cases – see spikes on charts – but in different cases
  - V8 Corrects one specific issue that had been troublesome in V7: handing of namespace exceptions, which were causing problems in Cocoon Unit
  - Serializer slower with undesired behavior on disable-output-escaping
- **XSLTC**
  - Required code changes to get it working
    - >65K method size
    - Patch submitted <http://issues.apache.org/jira/browse/XALANJ-2206>
  - XSLT changes were also required (not fully explored yet)
  - Once working, the clear winner on performance

# Testing caveats

---

- During multi-user testing, JVM behavior was unpredictable, with occasional long Full GCs (15+ sec)
  - These in turn impacted average response times as well as CPU utilization
  - We reran tests that had long Full GCs, but they highlight the fragility of both CPU and response time metrics
- Xalan 2.4 on JDK 1.4 failed the multi-user test
  - Unable to handle the load, Full GC meltdown
  - Results listed are extrapolated from a partial run
- On the whole, however, we feel comfortable that the results provide a reasonably accurate measure of relative XSLT processor performance in Cocoon

# More details ... and Gregor

---

- More details and additional downloads are available here:
  - [www.agilepartners.com/blog](http://www.agilepartners.com/blog)
  - Including full details on the test and Cocoon configuration
- We plan to run the same tests on Gregor
  - Will update the article and notify dev list when available