On the Software Ecosystem Health of Open Source Content Management Systems

Sonny van Lingen[1], Adrien Palomba[1], Garm Lucassen[1]

[1]Utrecht University
{s.j.vanlingen, a.r.v.palomba, g.g.lucassen}@students.uu.nl

Abstract. Choosing a content management system on which you rely your business is challenging because they need a healthy software ecosystem in order to function efficaciously. Unawareness of this will result in content managers having uncertainty about the future suitability of their chosen content management system. This study describes an empirical, inductive approach by comparing the software ecosystem health of the three most popular open source Content Management System platforms (WordPress, Joomla and Drupal) according to a number of health characteristics. Taking the software ecosystem health of a desired content management system into account enables stakeholders to make a more grounded decision in choosing either of the Content Management Systems. This could lead to a more suitable, dynamic and/or sustainable solution.

Keywords: software ecosystems, software ecosystem health, Drupal, Joomla, WordPress, content management systems

1 Introduction

Using an online Content Management System (CMS) to create and add content to a dynamic website is increasingly growing popular [14]. Designing an attractive website with the help of an online CMS is done with much more ease than having to perform a manual hard-coding process. A large amount of CMS platforms are offering turnkey solutions; however, specific features are mostly not available in a basis CMS installation package. In this case the content manager (website administrator) resorts to additional plugins. Plugins are collections of files developed by a third party, adding functionality to the core of the CMS platform. Therefore, the CMS platforms and the responsible developers for writing third party modules are part of a software ecosystem. Software ecosystems are defined by Jansen [6] as 'a set of actors functioning as a unit and interacting with a shared market for software and services, together with the relationships among them. These relationships are frequently underpinned by a common technological platform or market and operate through the exchange of information, resources and artifacts.' Not being able to survive in a software ecosystem has already led to the demise of many software vendors [6]. Being a CMS platform, measuring the health of your own software ecosystem is essential. More so, for content managers who have to decide on implementing either one of the CMS platforms, this work can help in

making a sensible choice for either one of the CMS platforms (as the software ecosystem's health characteristics relate to its lifetime expectations). WordPress, Drupal and Joomla all act as software ecosystem orchestrators (in this context, as a vendor) by providing third party module developers the opportunity to develop plugins within an open platform. This has lead us to pursue the following research question:

What is the health of the Software Ecosystems of the three most popular open source content management systems?

It is necessary to understand that WordPress is responsible for a significantly higher market share *(53,6%)* than both Joomla *(9,6%)* and Drupal *(6,4%)* [18] and that WordPress' community of third party developers is not comparable to both the Drupal and Joomla ecosystem in terms of development maturity. It demonstrates that Drupal and Joomla are in a battle for the second spot in the open source CMS market, behind WordPress. Furthermore, it is necessary to understand that we consider content managers who are not involved in developing modules not to be active contributors in the ecosystem because they do not make an active contribution - they are solely using the system.

The practical contribution of this research is to provide detailed information that describes the software ecosystem health of the three CMS platforms at both a platform-level and a module-level. This is done by measuring a number of software ecosystem health characteristics, which are described elaborately in the research method section. This is done by a mixture of computational calculations, our own observations and a survey (to confirm the aforementioned findings). In gathering data and response for the survey we heavily relied on communities (forums) and plugin overviews on the official sites of the three platforms' websites. Communities and plugin overviews of unofficial, third party sites are not taken into account in this research as their reliability (and their completeness) could be questioned.

This section introduced the notion of software ecosystems, software ecosystem health and its relation to CMS platforms. The remainder of the paper is structured as follows: In Section 2 a literature study is described, which defines key terms and provides definitions, besides identifying studies that support our research topic. Section 3 details the research methodology that was applied. Section 4 presents an analysis of the results together with the findings of our research. Finally, the conclusions, the limitations and an outlook for further research are provided in Sections 5 and 6.

2 Related Literature

In order to comprehend and explain the context of our study we carried out a focused literature review. In this section we consecutively describe three different aspects that have an important relationship with the software ecosystem health of CMS platforms, namely: software ecosystems, software ecosystem health and finally CMS platforms (and comparisons thereof).

It is observed that the notion of software ecosystems is still remarkably young; the first definitions are coined in the year 2003. However, up until 2008 the concept of ecosystems in a software or information technology perspective was still considered "not directly obvious" [11]. Various definitions of the notion of software ecosystems exist [2, 6, 10, 11]. We however consider the definition of Jansen the most appropriate to this study, which can be found in detail at the introduction section of this work.

According to Jansen, Brinkkemper and Finkelstein software ecosystems can be one of the following types: (1) market, (2) technology, (3) platform and (4) firm [7]. Within each type there are a number of factors that can help in reducing the scope of the software ecosystem. This study can be placed in the third category; this study's goal is to compare the software ecosystem health of three CMS platforms. Jansen and Cusumano [8] provide a classification model for software ecosystems, which is applied to 19 cases previously explored in software ecosystem literature. Finally, Campbell and Ahmed propose an elaborated three-dimensional view on the software ecosystem model explained by three central pillars: business; architecture and social aspects [3]. Software ecosystem health indicators are part of a software ecosystem related survey carried out under representatives of the Dutch Software industry [1].

As early as 2003, McKeever recognized the shift from static, manually deployed web content to dynamic, automatically deployed web content and the potential of content management systems in this perspective [12]. The maturity of CMS's has grown due to new web technologies, plus the need for improved role based web management that has supported this growth [16]. This growth in maturity and popularity has resulted in the fact that ~31.7% of today's websites are managed by a CMS platform [17]. Some works already compared CMS platforms by using other, non-ecosystem-related metrics. In a Search Engine Optimization (SEO) comparison experiment of the Joomla, Drupal and Wordpress CMS platforms, Drupal came out as the platform generating the most search engine revenue (2099 unique visitors from search engines in six months), followed by Joomla (1619 visitors) and WordPress (1439 visitors) [15].

A performance analysis of CMS platforms, again comparing Joomla, Drupal and Wordpress, reveals that the Joomla platform is best suited for novice content managers, whereas Drupal is suited better for content managers having to perform critical tasks and having to provide an increased flexibility [9]. A security audit report detailing the technical security of the Joomla and Drupal platform revealed unpleasant results; as of August 2009 the platforms were considerably safe but both platforms possessed a number of threatening security malfunctions [13]. Although it is not formally confirmed by another research engagement that these security malfunctions are not to be seen anymore, it is more than likely that these security threats are fixed at this moment in time.

3 Research Method

Reviewing the software ecosystem health of the three CMS platforms has led us to decide on a number of software ecosystem health metrics, partially inspired by economic ecosystem health characteristics [5] namely: (1) niche creation, (2) productivity and (3) robustness. A number of these health metrics are computationally measured,

which puts us in the position to process large amounts of data which would otherwise be impossible to review. Furthermore, a number of health metrics are measured manually. Finally, to confirm our findings, we carried out a brief survey under members (website administrators, module developers, core developers) of the ecosystems of interest, researching how they perceive the ecosystem health of the platform of choice. To this end we retrieved a random sample of respondents of interest. This sample consists of members of the three platform's community forums, workshop participants[1] and the authors' professional relations. The complete list of health metrics looks as follows:

- Growth of the platform (computational)
- Identification of the contributors (computational)
    - Including the number of unique developers.
- Up-to-dateness of modules (computational)
- Findability of the ecosystem (computational)
- Centrality of the platform (manual)
- Market share analysis (manual)
- Level of contribution per community user (manual)
- Perceived ecosystem health (survey, manual)

In order to perform the computations needed for the computationally measured health metrics we have developed a set of tools using either the PHP or Java platform. All of these tools exploit the mechanism of HTML parsing, which consists of browsing the HTML code of a given page to seek for a given value, since neither WordPress's, Joomla's nor Drupal's platform offer an Application Programming Interface (API) for executing search queries.

All of these programs have been executed from servers within the Netherlands, all using exclusively Dutch IP addresses[2]. During one encounter we faced a call limit per IP address. This has been solved by resorting to a VPN service which allows changing the external IP address on set intervals. A pool of exclusively Dutch IP addresses has been used for this purpose. The data gathering process started on 28 December 2012 and ended on 3 January 2013. Data originating from the year 2013 is filtered as we are only taking entries up to 31 December 2012 into account during the analysis. This has been decided to assure the analysis has a consistent end-date for all three platforms. We retrieved two collections of data:

- All official extensions for WordPress, Drupal and Joomla including every relevant field provided on its originating website (including name, author, date of creation and date of last modification).
- The number of Google hits per individual module.

The data utilized for measuring the manually measurable health metrics did not include computational interference - this data was accessible in a usable format right away.

---

[1] Participants in the 'Dutch Student Workshop on Software Ecosystems 2013'.

[2] Hereby avoiding retrieving different results given different geographical ranges of IP ranges.

4 Results and Data Analysis

This section presents our findings and the data analysis. These results are provided in table 1. Finally, to confirm our findings, we provide the results of the survey. These results are provided in table 5.

Table 1. Results overview per ecosystem health metric.

HEALTH METRIC	RESULTS
Growth of the platform in modules	**Drupal** has (and always has had) a larger number of modules for within its platform. Both **Joomla** and **Drupal** have shown a rapid growth after their respective introduction years. *Additionally visualized in figure 1.*
Growth of the platform in number of developers	The **Drupal** platform had always had a larger number of unique developers than **Joomla**, except for a limited period in 2010. During this period, **Drupal** failed to provide core updates for two years. *Additionally visualized in figure 2.*
Identification of the contributors	Total number of developers as of January 2013: **WordPress** *(9,904)* **Drupal** *(6,309)* **Joomla** *(3,360)*
	Average number of modules per developer as of January 2013: **Drupal** *(3.09)* **Joomla** *(3.01)* **WordPress** *(2.31)*
Up-to-dateness of modules	Percentage of modules updated in the year 2012: **Drupal** (59.62%) **Joomla** (41.57%) **Drupal including sandbox** (41.53%) **WordPress** (44.62%)
Findability of the ecosystem	**Joomla's** findability decreases from the year 2010 and on. **WordPress's** findability increases from that point. **Drupal** remains a smaller, niche player. *Additionally visualized in figure 3.* **Drupal** and **WordPress** show only a few cases of unfindable modules (~5%). **Joomla** suffers of approximately half its modules not generating results. **WordPress** seems to operationalize a slightly more effective SEO strategy. *Additionally elaborated upon in table 2.*
Centrality of the platform	**Drupal** is the most centralized platform, followed by **WordPress** and **Joomla**. *Additionally elaborated upon in table 3.*
Market share	**WordPress** possesses the largest market share (53,6%), **Joomla** (9.6%) and **Drupal** (6,4%) are battling for the second spot [18].
Level of contribution	**WordPress's** forum community possesses the largest number of topics and posts. *Additionally elaborated upon in table 4.*

A couple of remarks are to be made considering these results. Sandbox modules are modules which are not fully operational (yet). In the first two health metrics, WordPress could not be included as the platform does not publish the module's date of creation. The 2010 deviation, as can be seen in figure 2, could be explained by an ecosystem transfer of unsatisfied Drupal developers migrating to the Joomla ecosystem. During this period Drupal failed to provide core updates for two years whereas Joomla was releasing a major beta.

Currently, WordPress attracts more developers to join their ecosystem. On average however, the WordPress module developers are slightly less productive. In analyzing the up-to-dateness of modules, the effect of including sandbox modules in this analysis is remarkably. Considering sandbox modules, the up-to-dateness is equal over all three platforms. Furthermore, in analyzing the findability of the ecosystems, we made use of Google's Trends functionality to gain an insight in the platform's findability. In analyzing the findability of individual modules we resorted to a self-developed tool, performing Google searches based on module name. Because there is no consensus on the term module (plugin and extension are also often used), we included all three terms. Scores have been normalized in order to remove false positives (only including scores where *-3 < z-score >3*). The query used has the following form:

```
"module MODULE_NAME" OR "plugin MODULE_NAME" OR "extension
MODULE_NAME" + CMS_NAME"
```

In analyzing the centrality of the platforms, we based our findings on the platform's official communication channels. In analyzing the market shares of the platforms, an important remark has to be made: the number of weekly downloads of the platform's executable is declining vastly (WordPress -34,4%, Joomla -24,0%, Drupal -32,2%) [18]. This might imply that the open source CMS market has already matured. Finally, in analyzing the level of contribution, WordPress could not be included entirely. Their forum community does not publish detailed information per member.

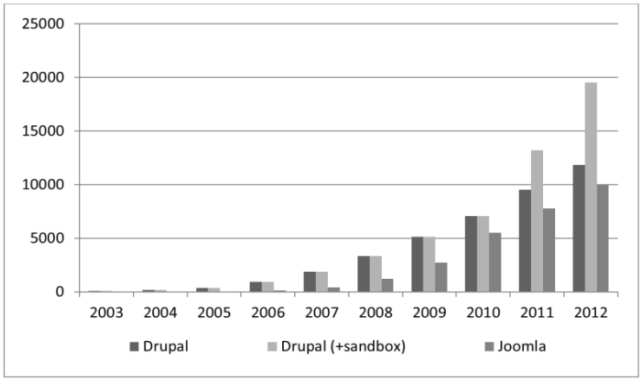

Fig. 1. Growth of the number of modules per CMS platform.

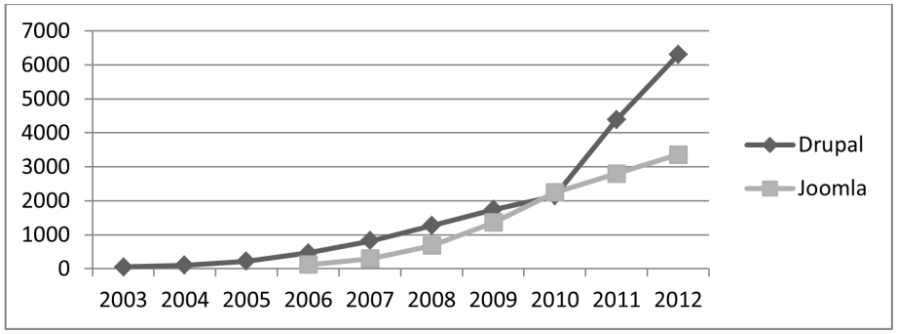

Fig. 2. Growth of the number of unique module developers per CMS platform.

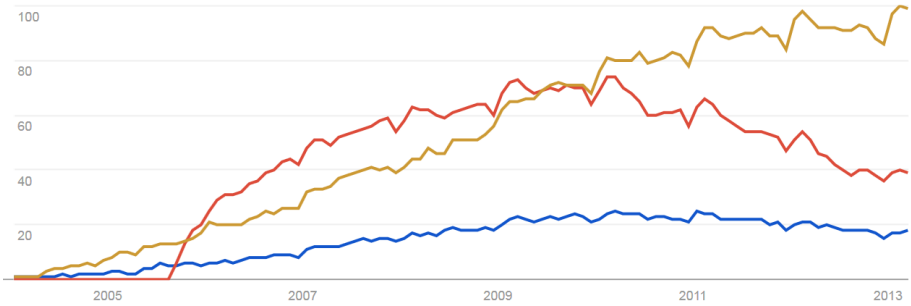

Fig. 3. Google Trends analysis of findability of the platform [4]. *From the bottom and upwards (at the commencing of 2013): Drupal, Joomla, WordPress.*

Table 2. Descriptive statistics of module findability on Google.

	Drupal	***Joomla***	***WordPress***
AVG	5,447	4,856	6,118
MAX	1,720,000	1,640,000	1,950,000
MIN	0	0	0
STD DEV	48,235	56,996	51,098

Table 3. Centrality of the platforms.

	Drupal	***Joomla***	***WordPress***
Module centrality	**Y**	N	**Y**
Support forum centrality	**Y**	**Y**	**Y**
Organized event centrality	**Y**	**Y**	**Y**
Documentation centrality	**Y**	**Y**	**Y**
Availability of support (platform-level)	**Y**	N	N
Availability of support (module-level)	**Y**	N	**Y**

Table 4. Descriptive statistics of the platform's forum communities.

	Drupal	***Joomla***	***WordPress***
Total number of members	1,334,960	593,517	*irretrievable*
Total number of topics	301,098	647,853	913,912
Total number of posts	1,115,507	2,718,144	3,090,335
Average number of posts per member	0.8357	4.580	*incalcuble*
Average number of topics per member	0.2256	1.094	*incalcuble*

4.1 Perceived Ecosystem Health

In the previous subsections we described and measured a number of ecosystem health metrics. The outcome of these measurements serves as a factual, raw data dependent mean to measure the given ecosystems' health. In order to compare these findings to how a number of stakeholders *(n=23)* perceive the ecosystem's health, we carried out a brief survey. Beforehand, it is hypothesized that a substantial amount of respondents regard the ecosystem health of their platform of choice equivalently to the findings presented previously. Furthermore, it is hypothesized that respondents also identify themselves with the poor prospects of the 'traditional' notion/operationalization of CMS platform providers and the upcoming shift to SaaS CMS solutions. The survey's outcome is described in table 5. Note that some comments were rephrased because they were either in Dutch or otherwise in a format not fit for citing. We tried to apply this rephrasing as sharp and precise as possible. We have removed entries from respondents who did not complete the survey in a professional, plausible way.

A couple of noteworthy comments were given. One respondent is unsure of our chosen cloud naming convention, stating that cloud based CMS solutions should be referred to as PaaS. We however do not second this vision – such cloud based solutions do seldom offer an actual platform. Furthermore, one respondent is using WordPress and Joomla and considers WordPress to be much easier for end-users than Joomla. Finally, one respondent is reluctant to embrace SaaS CMS solutions. Additional cost is not seen as the major drawback:

"I always want to host my website and data myself. I am absolutely not comfortable with SaaS suppliers being able to access my (personal) websites and data. "

A couple of remarks are to be made considering these results. In the first question, the distribution of the results is similar to the platform's respective market shares. In the second question, respondents could select more than one checkbox so these numbers, in total, exceeds the number of respondents.

One respondent is already actively migrating websites to other platforms, whereas another respondent feels that WordPress needs to revise their strategic decision about the (lack of) templates.

Table 5. Summarized representation of the survey's outcome (r = number of respondents).

QUESTION	RESULTS
1. Which CMS solution are you currently most involved with?	WordPress (r=10) Joomla (r=7) Drupal (r=6)
2. In which way are you currently involved in the previously selected CMS platform?	Content Manager (r=22) Module developer (r=6) Platform core developer (r=3)
3. Are you worried about the (future) well-being of the previously selected CMS platform?	Not at all, complete trust (r=15) I have reasonable doubt (r=7) I will drop the platform as soon as possible (r=1; Drupal)
4. Have you heard about cloud computing SaaS (Software as a Service) CMS solutions?	Yes, I have (r=11) No, I have not (r=12)
5. Are you currently planning on migrating to another CMS solution?	Yes, I am (r=3) No, I am not (r=20)
6. Most SaaS CMS solutions are paid services. Assuming they suit your demands better, would you consider migrating to them despite the additional cost?	No, I will not consider paying (r=10) Yes, paying for a services does not bother me (r=5) Maybe, I will first need to have more information (r=8)

When asked for SaaS CMS solutions already known to the respondents, a number of solutions were named explicitly:

- WordPress (4 times);
- Netfirms.com;
- Shopify;
- Silkapp;
- Square Space;
- Google Sites;
- TransIP;
- LightCMS;
- WIX.

Three respondents declared they are currently planning on migrating to another CMS solution. When asked for clarification, the first mover declared to be migrating to the Drupal platform. The second mover declared to be moving to an unnamed SaaS CMS environment. The third and last mover declared to be moving to a self-developed CMS.

Finally, a substantial amount of respondents does not feel informed sufficiently about the potential of SaaS CMS solutions (*r=8* out of a sample of *n=23*).

5 Discussion

For this research to become more mature and to allow for a more powerful comparison of the platforms, future research could be devoted to retrieve and analyze more historical data about the platforms and its modules. Even though the data set used for this research was large and detailed, we encountered some limitations within this research. Therefore completeness is not claimed.

Firstly, we could only resort to publicly available data. In spite of the fact that this allowed for a rich 'snapshot' of the software ecosystem's health during the period in which the data was gathered, we had limited access to historical data. In the context of this research, more historical data would have proven to be useful.

Secondly, a number of other software ecosystem health characteristics are not elaborated upon. This is, on one hand, related to restrictions of the data available. The most apparent deficiencies of the data are the lack of a comparable number of downloads per module for the platforms and the lacking possibility to download (and analyze) modules computationally (thus, automatically). Due to the fact that Joomla decentralized the hosting of modules we were unable to retrieve these modules computationally, disqualifying them for automated code analysis. On the other hand, the chosen health characteristics, metrics and its subsets are based on the authors' intuition and expertise. These characteristics do not necessarily follow theoretical classifications and considerations in the soundest way, which might have resulted in missed opportunities in the selection and/or operationalization of health characteristics.

Thirdly, it is to say that the number of Google hits representing a particular module could be questionable, as it might have triggered an unknown number of false positive results. Normalization of the results still does not guarantee that we succeeded in excluding all false positives. However, this eliminated a large part of the outliers.

Finally a remark is to be made about the platform's end-user base. A large number of WordPress' SaaS-users are using the platform as a (personal) blogging tool – opposed to a relatively larger number of professional appliances by their competitors. Due to feasibility reasons these differences in 'content-maturity' have not been analyzed.

6 Conclusion

The main goal of this paper was to measure and compare the software ecosystem health of the Drupal, Joomla and WordPress CMS platforms. This has been done by empirically measuring a number of health metrics, for which we computationally and manually retrieved data. The focus of this comparison was at a platform-level and a module-level.

The results show that the Joomla and Drupal platforms have a comparable market share. Both market shares are significantly smaller than that of the market leader, WordPress. Furthermore, the results show that Drupal's level of growth has exceeded

Joomla's level of growth. Unfortunately, we were unable to retrieve comparable data for the WordPress platform, which would have enabled us to elaborate on the growth of this platform's number of modules and unique developers. Next to this, it is observed that the full-project modules within Drupal's platform are more up-to-date than Joomla's and WordPress' modules (that is, excluding Drupal's sandbox modules). Including these sandbox modules makes the three platforms' up-to-dateness surprisingly equal. Finally, it is observed that Drupal's platform is more centralized than Joomla's and WordPress platform.

Despite the fact that not all metrics are in favor of Drupal's platform, we conclude that Drupal's platform possesses a healthier ecosystem. Hereby it is taken into account that the results for the WordPress platform could not be properly supplemented to two of the health characteristics. These results lead us to conclude that the criteria used by the CMS users to choose a CMS are not primarily based on the health of its ecosystem. Furthermore, given our investigation on Google hits for modules and the platform as a whole, neither of these criteria seem to be a nontrivial criterion for users.

That said a few remarks are to be made. Firstly, the most recent Google's Trends analysis shows a slight downward trend for Drupal's and Joomla's CMS platforms. This suggests that both platforms already have matured and might lose a factor of their popularity in coming years. Furthermore, it is to be noted that the average number of weekly downloads declined vastly for these platforms. This implies that both ecosystems are in the process of becoming unhealthier, or that the open source CMS market is experiencing a (temporary) loss of popularity. However, this does not affect the WordPress platform.

To summarize: based on this research, Drupal's platform is the healthier one of the three platforms, despite of being the least popular. The results of the survey give to think that SaaS CMS solutions have not yet become a threat to "classical" CMS's. Solutions of this kind will probably mature in the future and will require new investigations to quantify its evolution.

References

1. Van Angeren, J., Blijleven, V. and Jansen, S.: Relationship Intimacy in Software Ecosystems: A Survey of the Dutch Software Industry. Proceedings of the Conference on Management of Emergent Digital Ecosystems (MEDES 2011) .
2. J. Bosch. From software product lines to software ecosystems. In Proceedings of the 13th International Conference on Software Product Lines (SPLC) . Springer LNCS, 2009.
3. Campbell, P.R.J., Ahmed, F.: A Three-Dimensional View of Software Ecosystems. Proceedings of the Fourth European Conference on Software Architecture: Companion Volume (ECSA '10). 81–84 (2010).
4. Google: Google Trends, http://www.google.nl/trends/explore#q=drupal, joomla,wordpress, (2013).

5. Iansiti, M., Levien, R.: The Keystone Advantage: What the New Dynamics of Business Ecosystems Mean for Strategy, Innovation, and Sustainability. Harvard Business Review. 20, 2, 1 (2004).
6. S. Jansen, A. Finkelstein, and S. Brinkkemper. A sense of community: A research agenda for software ecosystems. 31st International Conference on Software Engineering, New and Emerging Research Track , pages 187–190, 2009.
7. Jansen, S., Finkelstein, A., and Brinkkemper, S. Business network management as a survival strategy: A tale of two software ecosystems. In Proceedings of the First Workshop on Software Ecosystems. CEUR-WS, vol. 505, (2009)
8. Jansen, S., Cusumano, M.: Defining Software Ecosystems: A Survey of Software Platforms and Business Network Governance. Proceedings of the international Workshop on Software Ecosystems 2012. 1–18 (2012).
9. K Patel Savan Rathod V R Prajapati, J.B.: Performance Analysis of Content Management Systems- Joomla, Drupal and WordPress. International Journal of Computer Applications 0975 – 8887. 21, No.4, (2011).
10. Kittlaus, H.-B., Clough, P.N.: Software Product Management and Pricing: Key Success Factors for Software Organizations. Springer (2009).
11. Kuehnel, A.-K.: Microsoft, Open Source and the software ecosystem: of predators and prey—the leopard can change its spots. Information Communications Technology Law. 17, 2, 107–124 (2008).
12. McKeever, S.: Understanding Web content management systems: evolution, lifecycle and market. Industrial Management Data Systems. 103, 9, 686–692 (2003).
13. Meike, M. et al.: Security in Open Source Web Content Management Systems. (2009).
14. Mooney, S.D., Baenziger, P.H.: Extensible open source content management systems and frameworks: a solution for many needs of a bioinformatics group. Briefings in Bioinformatics. 9, 1, 69–74 (2008).
15. K. Patel, Savan; A. Patel, Jayesh; V. Patel Amit. International Journal of Computer Applications, vol. 52, issue 3, pp. 1-5.
16. Raghavan, N., Ravikumar, S.: Content Management System. 1–18 (2008).
17. W3Techs: Usage of content management systems for websites, http://w3techs.com/technologies/overview/content_management/all (2013).
18. Water&Stone: 2011 Open Source CMS Market Share Report. (2011).