

A Framework to Evaluate Open Source Software (EvalOSS)

Bhim Upadhyaya

bhim.upadhyaya@west.cmu.edu

<https://www.linkedin.com/in/bhim-upadhyaya-0648a223>

Integrated Innovation Institute – Center for Open Source Investigation
Carnegie Mellon University, Silicon Valley

NASA Ames Research Park, Moffett Field, CA 94035, USA
May 2, 2017





Agenda

Background

Technology Landscape

Business Landscape

Social Landscape

Analysis

Design

Architecture and Computation Design

Ranking Algorithm

Application of Ranking Algorithm

Conclusion

Future Work

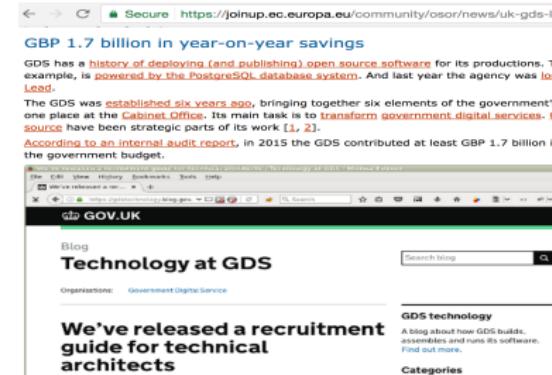
Background

Claims

- ① OSS is an important aspect of product suites.
- ② OSS saves huge costs.
- ③ OSS is inevitable for organizations.

Evidence

- ① Earlier research work [1][2][3].
- ② See UK government digital service (GDS) data.
- ③ See data and images.



Secure https://joinup.ec.europa.eu/community/osor/news/uk-gds-1

GBP 1.7 billion in year-on-year savings

GDS has a history of deploying (and publishing) open source software for its production. An example, is powered by the PostgreSQL database system. And last year the agency was led.

The GDS was established six years ago, bringing together six elements of the government's one place at the Cabinet Office. Its main task is to transform government digital services. GDS have been strategic parts of its work [1, 2].

According to an internal audit report, in 2015 the GDS contributed at least GBP 1.7 billion to the government budget.

Technology at GDS

We've released a recruitment guide for technical architects

GDS technology A blog about how GDS builds, finds and runs its software. Find out more.

Categories



www.pcmag.com/article2/0,2817,2367829,00.asp

PC REVIEWS BEST PICKS HOWTO NEWS TIPS BUSINESS EXP

#PCMagPicks #BoseSpying #GoogleHome #GalaxyS8 #FacebookFB

The Samsung Galaxy S8 on the best unlimited plan.

verizon /

Home / Business / Survey: 98 Percent of Companies Use Open-Source, 29 Percent Contribute Back

Survey: 98 Percent of Companies Use Open-Source, 29 Percent Contribute Back

BY DAVID MURPHY AUGUST 15, 2010 06:19PM EST • 0 COMMENTS

A recent survey by Zenoss shows that 98 percent of all enterprise companies use open-source software within their organizations. Yet a survey by Accenture shows that only 29 percent



Background contd...

Secure | <https://www.data.gov/open-gov/>

DATA.GOV

DATA TOPIC

OPEN GOVERNMENT

Data.gov is the federal government's open data site, and aims to make data increases citizen participation in government, creates opportunity both the private and public sectors.

Data.gov Implements the [Executive Order](#) on making government [Policy](#).

Open Data in the United States

Numerous states, cities and counties have launched open data site

- [Cities.Data.gov](#)
- [Counties.Data.gov](#)
- [States.Data.gov](#)

Secure | <https://unite.un.org/opensource>

UNITED NATIONS

Office of Information and Communications Technology

Welcome to the United Nations

App Quicklinks Resources Quicklinks Our Products News Events Multimedia

Our Products / Unite Open Source

Technology is always advancing the ways in which organizations relevant and cutting-edge innovation work of the United Nations.

Mr. Ameer Baig, Assistant Secretary-General for Economic Development

The United Nations Open Source Innovation Initiative (Unite Open Source) aims to break down barriers to technology innovation through open source governance, communities and collaboration.

OSS, Open Data, Open Government

Secure | <https://www.data.gov/developers/open-source>

DATA.GOV

DATA TOPICS IMPACT API

DEVELOPERS

Open Source Data Harvesting APIs Challenges

Open Source

Data.gov is powered by two open source applications, CKAN and WordPress, and it is developed by the Federal Government. CKAN provides a bird's eye view of all of the open source work currently being done. Beyond Data.gov, the Federal Government is part of a flourishing open source ecosystem government open source projects and open source community research.

Contribute to Data.gov on GitHub

In the spirit of open source software, everyone is encouraged to help improve this project.

Secure | <https://www.data.gov/open-gov/>

- Cities.Data.gov
- International Open Data

By category

Analysts U.S. [List]

Item	Link	Type
Argentina	http://www.datos.gob.ar/	International
Australia	http://data.gov.au/	International
Austria	http://data.gv.at/	International
Bahrain	http://www.bahrain.bh/wps/portal/datab/	International
Belgium	http://data.gov.be/	International
Brazil	http://dados.gov.br/	International
Canada	http://open.canada.ca/en	International
Chile	http://datos.gob.cl/	International
China	http://govinfo.nlc.gov.cn/	International
Colombia	www.datos.gov.co	International
Costa Rica	http://datosabiertos.gob.go.cr	International
Denmark	http://digitaliser.dk/	International



Technology Landscape

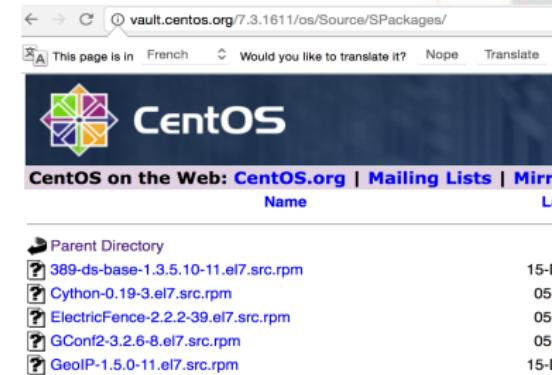
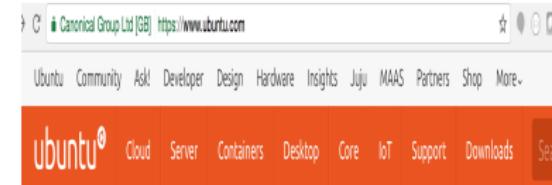
Sample Categories

- ① Operating System
- ② Office Automation
- ③ Programming Languages
- ④ Development Tools
- ⑤ Big Data Analysis



Technology Landscape - Operating System

- Less friendly source code crawling
- Appreciated more for desktop
- Also used for server
- Look and feel: Mac + Windows
- Source code crawling is better
- Close links with enterprise edition
- Close links with rapid development edition





Technology Landscape - Office Automation

OpenOffice

- Code structure
- Custom static code analysis

Mirror of Apache OpenOffice

5,747 commits 24 branches

Branch: **trunk** New pull request

Matthias Seidel Synchronized German translation (SDF<->Pootle) ·

File	Description
ext_libraries	Add initial support for building AOO with CI
ext_sources	Remove StAX API source tarball.
extras	Synchronized German translation (SDF<->Pootle)
main	Fixed typos (whitespace)
test	When otherwise absent, get the run and run
.gitignore	#i123623# RmMoz 8/9: remove the obsole

Office Automation

LibreOffice

- Different from OpenOffice code structure
- Need to write custom static code analysis

LibreOffice / core

406,907 commits 365 branches

Branch: **master** New pull request

j-carl committed with Michael Stahl tdf#39468 Translate German comm

File	Description
.git-hooks	git pre-commit hook: Also c
UnoControls	loplugin:checkunusedparar
accessibility	gbuild: Remove MSVC 2013

Technology Landscape - Programming Languages

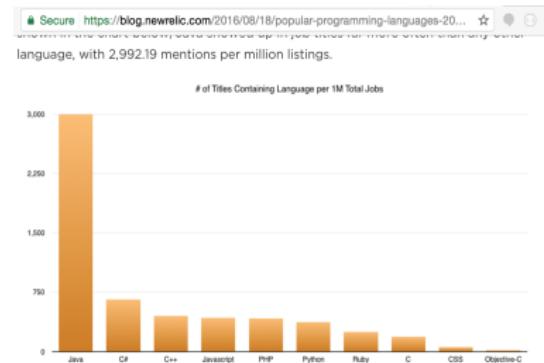
IEEE 2016 Rating

- Peer reviewed
- Scholarly
- More reliable

Topics	Reports	Blogs	Multimedia	Magazine
Language Rank	Types	Spectrum Ranking		
1. C	Code	100.0		
2. Java	Code, Diagram	98.1		
3. Python	Code, Diagram	98.0		
4. C++	Code, Diagram	95.9		
5. R	Diagram	87.9		
6. C#	Code, Diagram	86.7		
7. PHP	Code	82.8		
8. JavaScript	Code, Diagram	82.2		
9. Ruby	Code, Diagram	74.5		
10. Go	Code, Diagram	71.9		

New Relic 2016 Rating

- Commercial company
- Focused on job listing
- Important information but need external validation





Technology Landscape - Development Tools

KeyCDN

- Blogging
- Descriptive information
- intelliJ ↯ 4, Eclipse ↯ 5

The screenshot shows a web browser window with the URL <https://www.keycdn.com/blog/best-ide/>. The page lists the top 10 IDEs: Microsoft Visual Studio, NetBeans, PyCharm, IntelliJ IDEA, Eclipse, Code::Blocks, Aptana Studio 3, Komodo, RubyMine, and Xcode. Below the list is the heading "Finding the Best IDE for Your Needs".

G2Crowd

- Blogging
- Rating and descriptive
- intelliJ ↯ 4+ stars,
Eclipse ↯ 4 stars

The screenshot shows two G2 Crowd product pages. The first is for "Visual Studio" with a G2 Score of 92. It includes a logo, a star rating of 5 stars from 217 reviews, and a "ADD TO COMPARE" button. The second is for "IntelliJ IDEA" with a G2 Score of 79. It includes a logo, a star rating of 5 stars from 190 reviews, and a "ADD TO COMPARE" button. Both pages provide brief descriptions of the tools' capabilities.



Technology Landscape - Big Data Analysis

Big Data Frameworks

- Large number of choices
- Diversity in implementation technology
- Hadoop Vs. Spark

Code Base

[cloudera / hadoop-common](#)
forked from [ekoontz/hadoop-common](#)

Code Pull requests 2 Projects 0 Pulse Graphs

Branch: trunk → [hadoop-common / src / java / org / apache / hadoop / net /](#)

This branch is 153 commits ahead, 11 commits behind ekoontz:trunk.

Thomas White HADOOP-6926. SocketInputStream incorrectly implements read(). Contrib.

CachedDNSToSwitchMapping.java HADOOP-6688. Apply audience and stability :
 DNS.java HADOOP-6688. Apply audience and stability :
 DNSToSwitchMapping.java HADOOP-6688. Apply audience and stability :
 NetUtils.java HADOOP-6688. NetUtils:normalizeHostName
 NetworkTopology.java HADOOP-6884. Add LOG.isDebugEnabled() §
 NetworkTopology.java HADOOP-6884. Add LOG.isDebugEnabled() §

www.predictiveanalyticstoday.com/bigdata-platforms-bigdata-analytics-software/

50 Bigdata Platforms and Bigdata Analytics Software

IBM Bigdata Analytics, HP Bigdata , SAP Bigdata Analytics, Microsoft Bigdata, Oracle Bigdata Analytics, Talend Open Studio, Teradata Bigdata Analytics, SAS Big data, Dell Bigdata Analytics, HPCC System Big data, Palantir Bigdata, Pivotal Bigdata, Google BigQuery, Pentaho Big Data Analytics, Amazon Web Service, Cloudera Enterprise Bigdata, Hortonworks Data Platform, FICO Bigdata Analytics, Cisco Bigdata, Splunk Bigdata Analytics, Fusion-io Bigdata, Intel Bigdata, Mu Sigma Bigdata, MicroStrategy Bigdata , Opera Solutions Bigdata, Redhat Bigdata, Informatica Bigdata, MarkLogic Bigdata, VMware Bigdata, Syncsort Bigdata, SGI Bigdata, MongoDB , Guavus Bigdata, Alteryx Bigdata, 1010data Advanced Analytics, Actian Analytics Platform, MapR, Tableau Software bigdata, QlikView Bigdata, Attivio's Bigdata, DataStax Bigdata, Gooddata, Google Bigdata, Datameer, CSC Big Data Platform, Flytxt, Amdocs, Cisco Bigdata, Platfora, GE Bigdata

[apache / spark](#)
mirrored from [git.apache.org/spark.git](#)

Code Pull requests 478 Projects 0 Pulse

Branch: master → [spark / core / src / main / scala / org / apache](#)

asmith26 committed with srowen [MINOR] Issue: Change "slice" vs "par"

AsyncRDDActions.scala [SPARK-13928] Move org.apa
 BinaryFileRDD.scala [SPARK-16575][CORE] partitic
 BlockRDD.scala [SPARK-19998][BLOCK MAN/
 CartesianRDD.scala [SPARK-12692][BUILD][CORE]



Business Landscape

Sample Categories

- ① Consumer Electronics
- ② Managed Health Care
- ③ Retail
- ④ Manufacturing and Ag.
- ⑤ Online Payment

Business Landscape - Consumer Electronics

Open Eletronics

- OSS part of open electronics
- New challenge to evaluate



The screenshot shows the homepage of OpenElectronics.org. At the top, there's a navigation bar with links for BLOG, ABOUT, 3DRAG 3DPRINTER, PROJECTS, CONTEST, SUBMIT!, and CONTACTS. Below the header, the title "OPEN ELECTRONICS" is displayed in large, bold letters, with "SOURCE ELECTRONIC PROJECTS" underneath. A main feature section highlights an "An Open Source Sprinkler System powered by Yún". It includes an image of a circuit board and a small robot-like character. The text describes it as a control unit for a garden sprinkler that can be controlled via web and... Below this, there are other project cards for "An Electronic LED Roul Build", "An Open Source Sprinkler By Yún", "A Full Featured Mp3 D", and "Using A Telegram Bot To Experiment With H".

Consumer Electronics

OLPC

- Different problem to address
- Different strategy
- Different measure



The screenshot shows a browser window displaying the OLPC Wiki in Hindi. The URL is wiki.laptop.org/go/The_OLPC_Wiki/han-na. The page features a large orange number '1' and a green laptop icon. There are four colored arrows pointing outwards from the center. The page content includes sections for 'translate', 'page', 'discussion', 'view source', 'history', and 'help support oipo'. A sidebar on the left lists links for 'About OLPC', 'The OLPC Wiki', 'Contact us', 'Blog', 'Communicate', 'Participate', and 'laptop.org'. A footer at the bottom provides information about monitoring and language support.



Business Landscape - Managed Health Care

Connect

- Open source health information exchange
- Guide by HIPAA
- OSS influenced by HIPAA

The screenshot shows a web browser window with the URL www.connectopensource.org/about/what-is-connect. The page title is "About CONNECT". On the right, there's a photo of a man and a woman. The left sidebar has a yellow "About" tab selected, with other tabs for "What is CONNECT?", "FAQ", "Role in the Nationwide Health Information Network", "Promoting Meaningful Use", "Federal Agencies", "Project Milestones", and "Awards". The main content area describes CONNECT as an open source software solution for health information exchange.

Managed Health Care

OurMed

- Over 20 categories of OSS
- Over 100 different products

The screenshot shows a web browser window with the URL www.ourmed.org/wiki/List_of_open_source_healthcare_software. The page title is "List of open source healthcare software". It features a globe icon and the text "The Unbiased Source of Healthcare Information". Below the title is a navigation bar with links for HOME, ARTICLES, BLOGS, DISCUSSIONS, NEWS, EVENTS, ABOUT, and HELP. The main content area is titled "List of open source healthcare software".

Business Landscape - Retail

Keyhut

- Free POS system
- Can create multiple reports
- Can be run on a home computer too

Line	Pieces	Description	Stock Number	Price	Amount	Tax
1	5	COLOR KEY ONE SIDED	600008	2.99	14.95	0.750%
2	1	RUBBER GRIP DOMESTIC KEY	600088	4.29	4.29	0.750%
3	1	ALL ABOVE REDUCED 20%	-888888	-3.85	-3.85	0.000%
4	1	TWISTY KEY RING	60057	3.00	3.00	0.750%
5	1	1 UP MINUS	-777777	-0.50	-0.50	0.000%
6	5.250	BEADED KEY CHAIN / METER	67048	1.75	9.19	0.750%
7	1	COUPON / UNIT REDUCTION	-1	-5.00	-5.00	0.750%
8						
9						
10						

DATE: Description:

Inventory count: *****
PURCHASE #07680

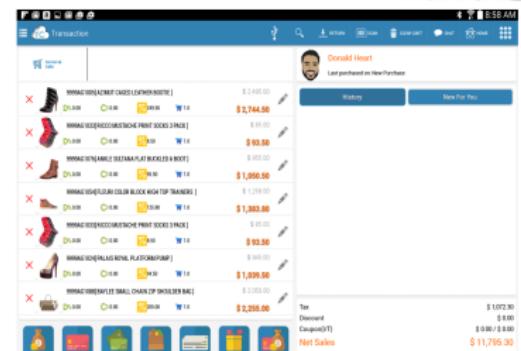
Enter Stock Number:	Price:	Quantity:
*****	0.00	0
	Tax Rate	

Press [+] to total.
F1 Stock table F3 Discount F5 Void F6 Park F7 Change Tax F9 Reduction F11 Back

Retail

ZeroPOS

- Free cloud-based POS
- Provides unlimited stores, registers, users and products



Business Landscape - Manufacturing and Ag.

OpenPro

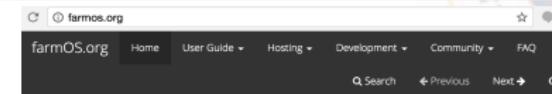
- Open source ERP system
- Modules: Manufacturing, Financials, Distribution, CRM, E-Commerce



Manufacturing and Ag.

farmOS.org

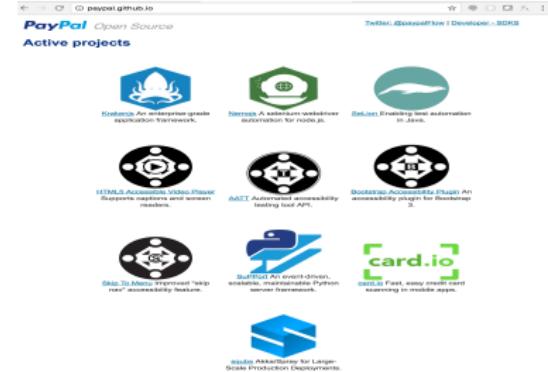
- Web application: Drupal
- Farm management, planning, and record keeping



Business Landscape - Online Payment

PayPal

- Wide variety of OSS projects
- Infrastructure is driven by OSS



Online Payment

cyclos

- Project of a network of Social TTrade Organization (STRO)
- Money movement





Social Landscape

- *Social media support*: Community based development requires networking and online interactions. A society supportive of social media is critical.
- *Government support for OSS*: At least government approval is required; could be in the form of free economics.
- *Investment model*: Individual and community based investment; donation model.
- *Regulations*: Should not restrict open development; regulations supporting free economic models are helpful
- *Openness*: Openness in manufacturing devices that support OSS has positive impact; this parameter is partially controllable.



Analysis

Observations

- ① Different projects have different priorities and hence the decisions.
- ② Criteria differ within the same company.
- ③ Business have fairly stable parameters compared to projects or departments.
- ④ Business parameters differ from one business to another.
- ⑤ Social parameters are the most stable.

Inferences

- ① There is no uniformity for priority parameters.
- ② Criteria cannot be fixed.
- ③ Might be able to use fixed categories.
- ④ Business parameters cannot be fixed. Fixed categories with varying weight might increase parameter stability.
- ⑤ Might be able to treat social parameters as fixed input.

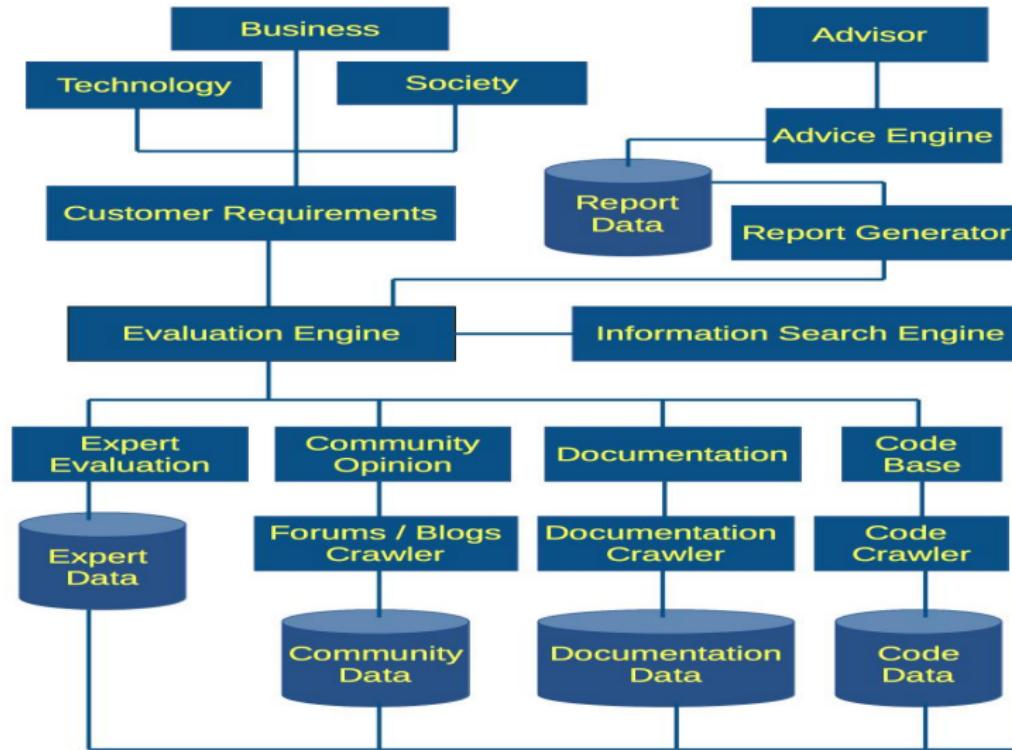


Design

Design Considerations

- A Exhaustive coverage of the information aspects
 - ① Expert evaluation
 - ② Community opinion
 - ③ Documentation
 - ④ Code base inference
- B Primarily a quantitative approach
- C Exhaustive coverage of requirement aspects – B , T , S
- D Scalability
- E Reconfiguration engines
 - ① Evaluation engine
 - ② Forums / Blogs crawlers
 - ③ Documentation crawlers
 - ④ Code crawlers
- F Customized report generation

Design - Architecture and Computation Design





Design - Ranking Algorithm

Ranking Algorithm

- ① Input value and weight for each input parameter:

$$R_{in} = (R_{value}, R_{weight})$$

- ② Populate the input set: $R = \{R_1, R_2, \dots, R_n\}$

- ③ Search and fetch all potential candidate products:

$$P = \{P_1, P_2, \dots, P_n\}$$

- ④ Compute impact of each product:

$$I(P) = \sum_{i=1}^k (\sum_{j=1}^m S_j * w_j) * R_{(i)weight}$$

- ⑤ Create an impact sequence: $I = [I(P_1), I(P_2), \dots, I(P_n)]$

- ⑥ Create a ranking map:

$$Z = [1 \mapsto I(P_h), 2 \mapsto I(P_s), \dots, n \mapsto I(P_t)]$$

- ⑦ Emit a selection or a selection set



Design - Application of Ranking Algorithm

			E	C	D	B	Intermediate I(P)	Rw	I(P)
			0.35	0.1	0.15	0.4			
	R1	Team skills	90	95	90	85	88.5	0.2	17.7
	R2	Features	85	90	85	85	85.5	0.25	21.375
Eclipse	R3	Cost of acquiring tool	100	100	100	100	100	0.15	15
	R4	Productivity	80	85	85	90	85.25	0.3	25.575
	R5	Reliability	90	85	95	90	90.25	0.1	9.025
			155.75	45.5	68.25	180			88.675
	R1	Team skills	80	90	85	90	85.75	0.2	17.15
	R2	Features	95	95	100	95	95.75	0.25	23.9375
IntelliJ	R3	Cost of acquiring tool	90	90	95	90	90.75	0.15	13.6125
	R4	Productivity	95	100	100	95	96.25	0.3	28.875
	R5	Reliability	90	95	95	90	91.25	0.3	27.375
			157.5	47	71.25	184			110.95



Conclusion

- Observed widely varying parameters across projects, departments, and businesses.
- *Non Uniformity ↳ Dynamism*
- Covered three main requirements aspects - B , T , S
- Covered exhaustive computational aspects.
- Easy to interpretation the results.
- Complex to implement.





Future Work

```
1a  com  equalinformation  evaloss  temp  scala  EvalOSSEngine.scala  EvalOSSEngine  VCS  VCS
2  com/equalinformation/evaloss/temp/scala/EvalOSSEngine.scala
3  /**
4   * Created by bpupadhyaya on 4/18/2017.
5   */
6  case class IntermediateRating( name: String,
7                                 expRating: Float,
8                                 commRating: Float,
9                                 docRating: Float,
10                                codeInfrating: Float
11                               )
12 case class Product( name: String,
13                      r1: IntermediateRating,
14                      r2: IntermediateRating,
15                      r3: IntermediateRating,
16                      r4: IntermediateRating,
17                      r5: IntermediateRating
18                     )
19 case class EvalParam(name: String, value: Float)
20 object EvalOSSEngine {
21   def main(args: Array[String]): Unit = {
22     val evalParams = Map( 1 -> "Team skills",
23                           2 -> "Features",
24                           3 -> "Cost of acquiring tool",
25                           4 -> "Productivity",
26                           5 -> "Reliability"
27                         )
28     val inputParams = Set(EvalParam(evalParams(1), 0.2f),
29                           EvalParam(evalParams(2), 0.25f),
30                           EvalParam(evalParams(3), 0.15f),
31                           EvalParam(evalParams(4), 0.3f),
32                           EvalParam(evalParams(5), 0.1f)
33                         )
34     val recomProducts = List( Product("Eclipse",
35                                         IntermediateRating(evalParams(1), 90, 95, 90, 85),
36                                         IntermediateRating(evalParams(2), 85, 90, 85, 85),
37                                         IntermediateRating(evalParams(3), 100, 100, 100, 100),
38                                         IntermediateRating(evalParams(4), 80, 85, 85, 90),
39                                       )
40   )
41 }
```



References

- [Wasserman et al., 2011] A. I. Wasserman, et al.
OSSpal: Finding and Evaluating Open Source Software, AFDA, 2011.
- [Stol and Babar, 2010] K. J. Stol and M. Ali Babar.
A Comparision Framework for Open Source Software Evaluation Methods, IFIP Advances in Information and Communication Technology, vol 319, 2010.
- OSSpal.
Available online at <http://osspal.org> as at April, 2017.
- Center for Medicare and Medicaid Services (CMMS), *jRAVEN*.
Available online at
<https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/nursinghomequalityinitis/nhqimds30technicalinformation.html> as at April, 2017.
- Microsoft.
Visual Studio Community Edition, Available online at <https://www.visualstudio.com/vs/community> as at April, 2017.
- Eclipse Foundation, <https://eclipse.org/org/> as at March, 2017.
- Red Hat Community, <http://community.redhat.com/> as at April, 2017.
- CentOS, <https://www.centos.org/> as at Feb, 2017.
- GitHub, <https://github.com/> as at April, 2017.



References contd...



OpenBRR.

Business Readiness Rating for Open Source. Available online at
http://www.immagic.com/eLibrary/ARCHIVES/GENERAL/CMU_US/C050728W.pdf as at April, 2017.



Wikipedia.

List of Linux distributions. Available online at
https://en.wikipedia.org/wiki/List_of_Linux_distributions as at April, 2017.



LibreOffice, <https://www.libreoffice.org/> as at March, 2017.



Apache OpenOffice, <https://openoffice.apache.org/> as at March, 2017.



[Grigore, 2017] R. Grigore.

Java Generics are Turing Complete. Available online at <https://arxiv.org/pdf/1605.05274.pdf> as at April, 2017.



[Madey and Freeh, 2017] G. Madey and V. Freeh.

The Open Source Software Development Phenomenon: An Analysis Based on Social Network Theory, Eighth Americas Conference on Information Systems. Available online at
http://www3.nd.edu/~oss/Papers/amcis_oss.pdf as at Feb, 2017.



[Shull et al., 2017] F. Shull, et al.

Open Systems Architecture: Progress and Challenges, Software Engineering Institute, Carnegie Mellon University. Available online at
http://resources.sei.cmu.edu/asset_files/Presentation/2015_017_001_447415.pdf as at April, 2017.