



International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Kaput Link Manager

Prof.Ajanthaa Lakkshmanan *, E.Aruna, A.Abirami Shri

School of Computing Science & Engineering,
VIT University, Vellore, India

Abstract— The appearances of fault links in the web are un-trustworthy and not reliable. The data which is removed from the website later detected as the dead links also called as kaput links .These dead links may be the problem to the internet users and leads to user dissatisfaction and time waste, the dead links in the web site’s as such as the e-commerce sites and business sites may leads to losses due to customer dissatisfaction. Recent surveys say that they are 10 % of dead links in the web. To overcome this problem we propose an approach that can detect the dead links and store the list of dead links in to a data base for future purpose it also separates external and internal links more over the internal links are more often verified.

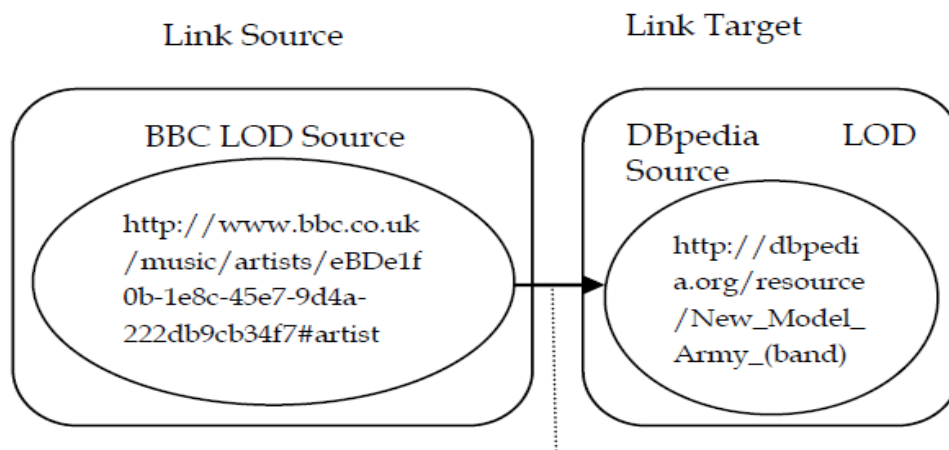
For example: www.childyart.com/laptops/Sony VPECH25EN/p/itmdc8cnbopg? pid = Com&ref=ad1-as4

The above link is related to the online E-commerce website and the product is deleted from the website but the hyper-link is not removed so when the customers interested to buy the product they access the hyper-link and they couldn’t find any information related to the product due to the dead link or kaput link this would be a serious problem to the customer and to the E-commerce website and the customer finds this incident as the unconcerned issue and stops buying in that E-commerce ,that leads to serious losses for the E-Commerce website and if they are more dead links in the other business then this leads to the loss of customers.

Keywords— Dead links, Kaput links, Anchor text, Extract Links, URL (Uniform resource locator)

I. INTRODUCTION

The internet consist of various Kaput Links or the “ Dead Links “ the setback of these Links is that there neither contains the information nor they provide the link to open .So these links are the sober problem to the internet users and to different technical people , that links provide the negative impression to the users of the different users and more problematic to the Ecommerce site because these site has to provide the quality of services to the user through online if there site have the dead or kaput links then this will affect their business and profit too .The Chinese Ecommerce named “chinghubu.org” in 1997 lost over 60,000 customers due to the dead-links in their vast site in the online-shopping sector and due that they incur over \$ 85,000 of losses. Even a software company in Japan named “Sanyo corporation” which deals with the different software products in 1999 due to the dead-links in their site one of their major client cancelled the 5-year contract, the client later states to the media that they feel it as the improper maintenance and they feel unconfident and cancelled their 5-year contract .This shows the sober impact of the kaput-links in the web in the real world. The Present System that allows the user to register and then it allows the user to enter the link as an input and the system in-turn checks the link as a dead-link or not and provides the various alternative links to the User. The System uses the two algorithms such as link extraction and the anchor detection, the link algorithm are useful for the extraction of the accurate link and uses the various search engines and the other algorithm reveals the anchor texts in the various sites and this is very useful in the extraction of accurate search.



Link owl: sameAs

Fig. 1. Simple Link between BBC Music and DBpedia data

II. DESCRIPTION

The user gives the web site address as the input then the system searching in the domain name server at the particular name and IP address then the particular name is available then it will be move to the further process or else then it searches the “INTER-NIC” records or it searches another “Registration Agency “for corresponding match of name lastly it shows the target.

III. SYSTEM ARCHITECTURE

The system takes the given link as the input from the user and then if the link is broken it is directed to the meta data that is verified for any match for previous results then it uses the search engine for the processing and then the Link Extraction algorithm and the Anchor text algorithm is used for the detection of the anchors texts and the list of the valid links are the alternatives for the given link.

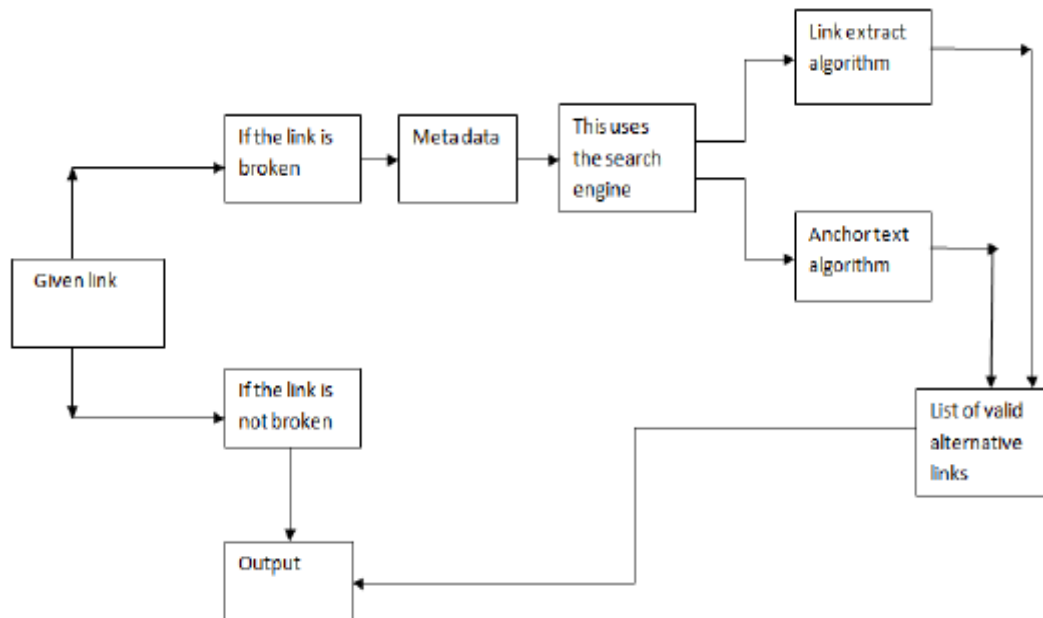


Fig. 2. System Architecture

IV. REQUIREMENT SPECIFICATION

The Requirements Specification document contains user requirements and analysis diagrams. Requirements are classified in to 2 ways:

A. Functional Requirements

The kaput Link handler can also be used in offline mode by searching over the hard-disk in the same way user can also check the links in a web site and except block of the particular website .The Kaput Links Handler will take the set of pages to find and a set of pages to weave and it manages the indirectly-linking the pages that will be checked and linked to the sub-directories also be checked.

B. Non-Functional Requirements

It contains of the Analysis and Design, Data Requirements, constraints, guidelines, and validation criteria.

1) Analysis and Design, Data Requirements

This phase will be consists of data flow diagrams, Data-dictionary .The data-Dictionary contains of statement of the process that shows the how the data was flows from the start to the end phase.

2) Constraints

These requirements do indirectly associated to the functions of the proposed system and it should be obligatory for the system that is developed. For detecting the Kaput Link it should first check the following:

- First the system searches in the Head, the website will be returned only header part not whole content.
- By using the Get, some of the websites do not returns the only headers and we can discard the header part arriving to the website.
- The system loads the WebPages after all other pages are being polled.
- The system rechecks the non responding web pages.

3) Guidelines

In the guidelines it checks the improvement of system usability. In this one, graphical user interface will decides it finds the variety of Links that system finds. The user can physically add different links that are to be fixed. The kaput link handler might think that some kaput links are fine that's why the system rechecks twice.

4) Validation Criteria

The validation criteria works in 2 ways they are internal and external.

a) Internal

Kaput link handler searches for relevant images and pages within the site and the system also checks for equal names in the website.

b) External

In the external mode the kaput link handler checks for the relevant pages in the given website externally. There are several names for a single name so the system uses different search engines to find the exact name and also uses the anchor text for the exact retrieval of the given user input.

V. ALGORITHM

```

Algorithm 5.1: ANCHORS(L)

comment: Extract inbound anchor texts from L.

H = {}
for each (ai, ui) ∈ L
  do {
    if ui ∈ H
    then Append(H[ui], ai)
    else H[ui] ← []
  }
return (H)
    
```

The above algorithm used for the extraction of inbound anchor text for the given user input the algorithm stores the given values H and it checks the name and the possible hyperlink and it relates to the value L, If the name s belong to the set in the H then it Appends the H with the u if not it changes set H to no value and returns the H.

```

Algorithm 5.2: ANCHORS(p)

comment: Extract inbound anchor texts for p

H = {}
S ← InAnchor(p)
for each url ui ∈ S
  {
    H[ui] ← []
    T ← Link(ui)
    for each page t ∈ T
  }
  do {
    {
      L ← ExtractLinks(t)
      for each (ak, qk) ∈ L
    }
    do {
      if qk = ui
      then Append(H[ui], ak)
    }
  }
return (H)
    
```

VI. EXISTING SYSTEM

The WebPages in the present world are more informative and due to the improvement of vast internet and technology and cheap internet the people can easily create more number of web pages but fails to maintain it in the future. Recent survey states that the 24 % of the people who create websites do not maintain it later. So the links are dead or broken in the different web pages .The existing system fails to detect dead links extensively and it do not check the hyper link and this leads to dead hyper links.

A. Advantages

The present system checks only the dead links and displays to the user and the accuracy of the search depends on the use of algorithms in the system and the precision of the data and speed depends on the use of Search Engines.

B. Disadvantages

The existing system is not accurate in searching the dead links and it is prone to more broken links and it doesn't uses the anchor texts and due to this the system is more prone to the dead links .

VII. PROPOSED SYSTEM

We propose a method that detects the broken links and also checks the hyper-Links of the websites and eradicates the broken links and it uses the hyperlinks and the anchors for the removal of the dead links. The system displays only the Valid Working links and the system also shows the other possible links for a given web page .The system uses the different searches for the exact retrieval of the information .It maintains a meta data that stores the previous search results and also stores in a referencing system this helps the browsers to redirect to the new location from the old location.

A. Advantages

The main advantage of the proposed system is that it also checks the anchors and the hyper-links that gives the accurate results of the given user input and the current system uses the meta-data that uses the previous results that can be

useful for the processing the future names if the same name is given again as an input .The current system uses the advanced search that uses the different search engines that gives the accurate results.

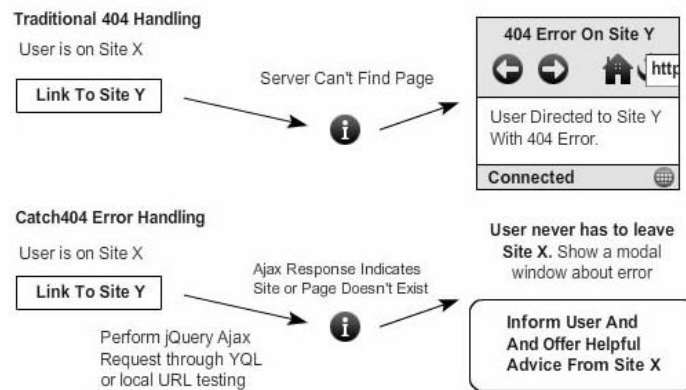


Fig. 3. Dead Link

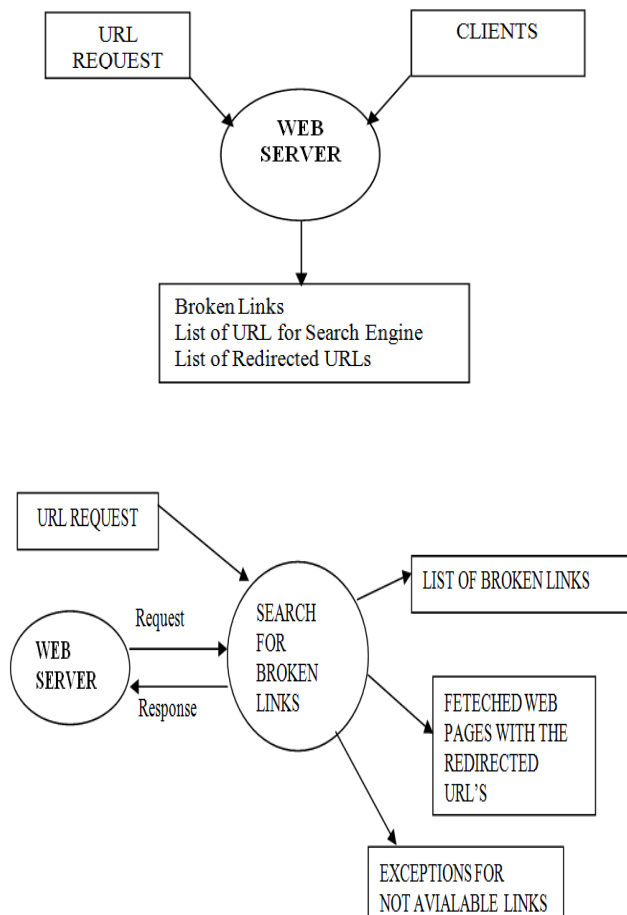
VIII. PURPOSE OF THE SYSTEM

The present system is mainly used for the dead link detection and also used for the accurate results and the system not only displays the dead links and Kaput Links it also displays the various alternatives good links to the user given input. This system integrates the meta-data that can be useful for speed up of the future process if user enter the same links again in the future .This system can be more accurate than the present system and can spot the broken links.

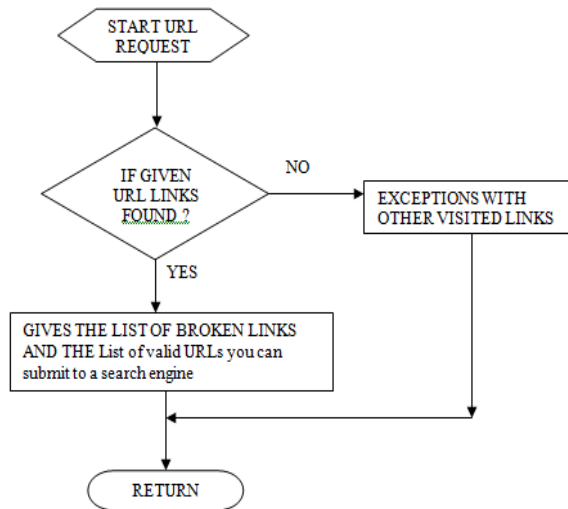
For Example:www.somecartinlive.org/Products/Tv/igonyprdfn_4646/?pid/dhjsdkj/ref=&sg

The stated is an online E-commerce site that sells different products but the product is removed from the list and the link is kaput link. If the user gives the above link the as an input but the system detects as the kaput link and provides the other choices of links such as the same product in different sites and the user can select any of the list of the links provided by the system.

IX. DATAFLOW DIAGRAM



FLOW CHART



X. CONCLUSIONS

We proposed a model that can not only the detection but also to provide the accurate alternative links the accuracy is depend on the search engines and given a URL the system detects whether it is a valid link or not and provides the alternative links for the given URL the using the Anchor texts algorithm and the Extract links algorithm and this will provide the valid links without the dead-links.

ACKNOWLEDGMENT

The authors would like to thank the School of Computing Science and Engineering, VIT University, for giving them the opportunity to carry out this project and also for providing them with the requisite resources and infrastructure for carrying out the research.

REFERENCES

- [1] S. Ameer, C. Bizzer, G. rraju, and Z. Ives, "Dbpeia: A nucleus in a web of open data," in Proceedings of the 6th International Semantic Web Conference (ISWC), ser. Lecture Notes in Computer Science, vol. 4825. Springer, 2010, pp. 722–754.
- [2] M. Malmate, "Making a web links Catalogue Part of the Semantic Web," in Proceedings of the International Conference on Dublin Core and Metadata Applications, J. Greenlace and W. Klter, Eds., September 2011.
- [3] Jan hu, C. hinder L. John, "A Broken link discovery framework for the web of data," in 2nd Linked Data on the Web Workshop (LODW2009), co-located with WWW 2009, Madrid, Spain, 2009.
- [4] Ram Iyer .J , "Hypertext link integrity," ACM Comput. Surv., vol. 31,no. 28, p. 28, December 2009.
- [5] Johan Paul, "Identity, Reference on the Web," in Proceedings of the Workshop on Identity, Meaning and the Web (IMW03), 2003.
- [6] P. Amer, J. Donald, H.Gober,"Triplify: light-weight linked data publication from relational databases," in WWW '08: Proceedings of the 18th international conference on World wide web. New York, NY, USA: ACM, 2008.
- [7] K. Andrewson, H. Roger, "The hyper-Link system," Journal of Universal Computer Science, vol. 1, no. 5, pp 210–254 2007.
- [8] S.Jing Lee, S.Henry Watt, "Fixing the "broken link" the w3objects approach," Computer Networks ISDN Syst., vol. 05, no. 8-10, pp. 1345–1350, 2009.
- [9] S. Salmam and P.Walt Mike, "A dead link framework for web," Int. D. Digital Library vol. 4, no. 1, pps. 849–578, 2008.
- [10] Y.G Ariman, "Broken names: handless in Web," Commun. ACM, vol.11, no. 4, p. 78, 2005. "