Advances in Differential Equations and Control Processes 2025, 32(2), 2343.
https://doi.org/10.59400/adecp2343

Article

Using learning and cognitive models for assessing quality of content hosted into websites

Shahida Tasneem Shaik[1], Hima Bindu Duddempudi[1], Kunal Sana[1], Kodanda Rama Sastry Jammalamadaka[1,*], Sasi Bhanu Jammalamadaka[2], Balakrishna Kamesh Duvvuri[3]

[1] Department of IoT, KLEF Deemed to be University, Vaddeswaram, Guntur District 522302, India
[2] Department of Computer Science and Engineering, CMR College of Engineering and Technology, Hyderabad 501401, India
[3] Department of Computer Science and Engineering, MLR Institute of Technology, Hyderabad 500043, India
*** Corresponding author:** Kodanda Rama Sastry Jammalamadaka, drsastry@kluniversity.in

CITATION

Shaik ST, Duddempudi HB, Sana K, et al. Using learning and cognitive models for assessing the quality of content hosted into websites. Advances in Differential Equations and Control Processes. 2025; 32(2): 2343.
https://doi.org/10.59400/adecp2343

ARTICLE INFO

Received: 21 December 2024
Accepted: 3 April 2025
Available online: 17 April 2025

Abstract: The quality of a WEB site is based on 42 different factors. Out of those factors, the quality of content is the key factor. Different types of content are hosted on websites, including text, data, text data, images, videos, graphics, and animations. There are also interrelations among the content sources. The element-level connectivity and hierarchical element connectivity of content sources dictate the quality of content hosted on the website. The number of content elements, connectivity among the elements and the number of hierarchies existing on a website are many and, therefore, require a framework using which the quality of content is computed. This paper presents a framework based on an expert and ANN model. Using the framework, the quality of content hosted on a website is computed considering all sub-factors determining the quality of the factor "Content". Using the framework proposed in this paper, the quality of the content was estimated accurately to 91.80.

Keywords: quality; assessment; computational methods; framework; WEB content; WEB sites

1. Introduction

Information exchange is predominantly taking place due to the advent of the Internet and WEB technologies. Most of the content these days is hosted on the WEB, making it easy for people to access the information in the least time and with negligible cost. The information posted on the WEB is in different forms, including text, images, audio, and videos. Information is also available in streamed mode.

Companies use websites for e-commerce and marketing. Web material is changing how people live, so its quality must be reviewed before being taken for granted.

Information seekers are changing their attitudes. Poor website quality will disappoint customers. If the website's content is inaccurate and irrelevant, it will disappear.

Usability, content, reliability, flexibility, functionality, portability, maintainability, privacy, security, information adequacy, safety, content, navigation, etc., can determine a website's quality. Website quality can be measured in various ways. Assessment can be done with evaluation tools. Many organizations develop video, audio, graphics, text, tables, forms, animation, logos, dynamic content, mouse-over effects, graphics, etc. Some sites are static, and some are dynamic. The literature contains many methods for computing the quality of WEB sites. Subjective assessments involve prejudice, and objective assessments lack completeness.

Some methods are biased by personal preferences or merely statistical observations of time, response time, etc. Stakeholders assess website quality differently. Programmers analyze a website's security, functionality, and maintainability, whereas consumers review its look and feel [1], structure [2], navigation [3], usability [4], multimedia [5], and completeness [6].

Each website has quality factors. These elements must be assessed to determine HOW well the WEB site meets user needs. Content plays a major role in dictating a website's quality. Many forms of content presentation exist, all associated with connectivity and structure. Connectedness and structure are the two major dimensions to consider when assessing a website's quality from a content perspective.

Complex websites make website quality assessment difficult. E-commerce, museum, and animation websites are complicated, and even website quality evaluation is more complicated.

Most website quality characteristics are interleaved and connected by complicated logic. Some quality characteristics are subjective, other's objective, and others quantitative. A composite model must include all quality computation aspects. A good quality evaluation model requires subjective, objective, and quantitative connections.

First, the expectations from a website must be determined. Quality parameters that relate to expectations must be determined and form a base for assessing the quality of a website. One of the factors included in the framework is content. The quality of content can be further recognized in terms of various sub-factors and each factor in terms of different characteristics. A framework is required to compute the quality of characteristics of each sub-factor, which are combined to compute the quality of factor "content."

The quality assessment methods required for computing the quality of characteristics of related sub-factors, the quality of sub-factors, and the quality of the factor "content" differ greatly.

Quality, as such, has relative behavior, and assessment can be graded on a scale. The quality of subfactors and characteristics falls into different quality grading levels. A cognitive model is required to compute the quality levels of different subfactors and characteristics. The quality assessment must be done much faster, pinpointing the specific area of weakness.

The number of computations to be made is huge, considering 42 factors, with each factor recognized as five subfactors and each subfactor having an average of 6 characteristics. The total complexity is around 1260 to be handled accurately and computed in real time, considering the average size of the website being around 100 pages of code through different web technologies.

While some sub-factors possess characteristics, others do not, necessitating direct quality computation using sub-factor or characteristic-level cognitive models. Learning models are required to predict the quality of certain subfactors and compute the quality directly through cognitive models related to these subfactors. The individually computed quality of different sub-factors must be combined to obtain the overall quality of each factor separately. Combining the quality of all factors will enable the computation of the overall website quality.

The content hosted on a website can be categorized into various types, including text, data presented in different forms, images, videos, graphics, animations, and audio. Various aspects of the content reflect its quality, which includes relevance, structure, information richness, and credibility; out of these, relevance and structure are the most critical issues. The content, which includes images, videos, graphics, animations, and audio, is generally grouped under the factor "Multimedia", leaving the issues of text, data, and structure to be mainly considered when assessing the quality of the content hosted on websites. The text, data, and their related structure mostly reflect the quality of the content, which is the focus of this research.

1.1. Problem definition

Thus, the quality of content must be computed using cognitive models at either the sub-factor or characteristic level and then predicted using machine learning models. A method must be used to compute the quality of the factor "content", considering the quality of its sub-factors, which are computed through different methods. Parsers are also required to parse code developed using various web technologies. A comprehensive framework must support all elements of the work.

1.2. Contributions of this paper

This paper presents a framework that includes cognitive models for assessing the relative grade of subfactor characteristics and an ANN-based learning model for predicting the quality of a subfactor based on the counts generated by parsers while processing the code related to the websites. There is also a need to find and develop parsers that compute the counts relating to the characteristics of website code. Counts provide characteristic values based on which the quality grades can be computed using the cognitive models and the quality predicted through ANN models.

Research questions: The following research questions are answered by this paper

1) How is the content of a website recognized?
2) How is the quality of content hosted on a website computed?
3) What are cognitive models, and how can they be used to assess the quality of sub-factors or their characteristics?
4) Will the machine learning models help assess the quality of characteristics of some of the sub-factors?
5) How is the quality of several sub-factors computed through different means combined to arrive at the overall quality of a factor?

1.3. Research outcomes

1) An algorithm (parser) that computes the web content connectivity and hierarchical connectivity.
2) A cognitive model that computes the quality of content in terms of its sub-factors and characteristics of the sub-factors.
3) An example set that can be used to learn an NN model for predicting the quality of a website's content.
4) A neural network model that can be used to predict the quality of sub-factors of the factor "content."

2. Literature survey

Ghalut et al. [7] have presented a method for computing the quality of video content rendered over wireless networks. They have used cluster analysis to assess the quality of video content based on random neural networks (RNN).

Lida et al. [8] have presented an editor for analyzing the value of text as a website's main structural component. They have emphasized the importance of the specialized editor for preparing quality website content. They have introduced a methodology called Universal Star (WQEMUS) with a theoretical and empirical basis for evaluating the quality of the website.

Sandra et al. [9] have proposed a website content auditing framework and combined it with content quality indicators. The authors used Parse Hub web scraping to get all the website details.

Ricca et al. [10] evaluated content, design, organization, and user-friendliness while assessing a website's quality. A website's organization includes page identification and hierarchical linking. The WEB pages are linked for easy navigation. Websites must be simple and user-friendly, presenting content according to user preferences.

Alwahaishi et al. [11] believe playfulness and content representation are the most essential aspects of website quality. Most basement-quality framework presentations lack a framework and computational approaches for website quality.

Hasan and Abuelrub [12] have proposed a general criterion for evaluating the quality of any website regardless of the type of service it offers. They contend that the quality criteria include content, design, organization, and user-friendliness. These dimensions, together with their comprehensive indicators and checklist, can be used by web designers and developers to create quality websites to improve the electronic service and the image of any organization on the Internet.

Singh et al. [13] found that the rapid proliferation of online apps necessitates quantitative evaluation. The objective evaluation of web applications is done using WebQEM. Web attribute weighting is subjective and primarily based on expert judgements. The authors developed a quantitative method for evaluating websites and applications. Their method helps evaluate product quality-affecting features, sub-characteristics, and attributes. They introduced website quality assessment models, methods, procedures, and principles.

Siddikjon et al. [14] have presented a novel approach to website assessment by adapting a methodology called WQEMUS (website quality evaluation methodology) based on a theoretical and empirical basis. They have used grounded theory to enable relevant concepts to emerge from data. They have presented a universal methodology that does not focus on a specific factor.

Alejandro et al. [15] have identified the main characteristics, methods, techniques, and tools and presented a multipurpose model that can be used to compute the quality of websites generically, considering every factor. No specific emphasis is given to the attributes' characteristics.

Hossein et al. [16] emphasized the importance of evaluating website quality, including content assessment, to identify strengths and weaknesses. They highlight that regular reviews can enhance user satisfaction and improve the overall effectiveness of websites, particularly in the mobile repair domain.

Luiz et al. [17] presented an automated approach for assessing the quality of collaboratively created web content using soft multi-view generation. This approach eliminates manual intervention by clustering correlated features, thus enhancing scalability and reducing classification error by up to 20%.

Raghavarao et al. [18] presented that assessing the quality of content on websites involves evaluating connectedness, hierarchical presentation, and various quality characteristics such as usability, reliability, and adequacy of information. Computational methods quantify these factors for a comprehensive quality assessment model.

Jessica et al. [19] identified the key quality characteristics of informative websites. They propose a technique for practitioners to enhance content quality through user surveys and web analytics, ultimately improving accessibility and reach to diverse audiences.

Alejandro et al. [20] highlighted over a hundred studies focused on assessing the quality of content on health-related websites, emphasizing their specificity and high citation rates. This indicates a significant academic interest in evaluating and improving website content quality.

Rim et al. [21] presented a method for assessing website quality, which includes evaluating content relevance, interactivity, design, and user feedback. Learnability and support for interactive content are crucial for educational sites, while social media sites focus on video quality and reputation criteria.

Alejandro et al. [22] presented various methods and tools for evaluating website quality, emphasizing the growing interest in this field, particularly in education, health, and commerce. They also highlight the importance of usability, information architecture, and user experience in content assessment.

Zhang et al. [23] presented a content quality assessment method that evaluates content based on user assessments, filtering keywords, and credibility scores. This method improves the accuracy of quality evaluations for web-hosted content through systematic analysis.

Li et al. [24] proposed a quality assessment method for web documents using random forest, focusing on content, structure, and access features. It includes a topic coverage degree calculation model based on LDA to evaluate the quality of web content effectively.

Pat et al. [25] assessed content quality using accuracy, timeliness, and relevance indicators. It found that while participants viewed websites as generally reliable and efficient, content quality was rated as moderate, highlighting the need for improvement in usability and functionality.

Daniel et al. [26] assessed the quality of user-generated content on collaborative websites using a feature selection method based on the SPEA2 multi-objective genetic algorithm. This method reduces the number of quality indicators while maintaining comparable error rates to existing methods.

Janusz et al. [27] focused on content-agnostic web browsing quality assessment methods. These methods evaluate quality without being influenced by the specific content or design of web pages, thus providing a more universal approach to assessing user-perceived quality.

Jayanthi et al. [28] used the Extreme Learning Machine (ELM), Husang et al. used SVM and the extreme learning method [29], and they also used extreme learning with regression [30] to assess the quality of the content hosted on the website based on the attributes selected by the user.

Research gap

While many have attempted to compute the quality of content from different dimensions, they have yet to consider the content's connectedness and hierarchical structuring. None have computed the counts of characteristics or the sub-factor using the parsers, based on which the quality is computed using either the cognitive or machine learning models. No framework has been presented for computing the quality of the content.

3. Methodology

The methodology used to evaluate the quality of content hosted on websites is illustrated in **Figure 1**.

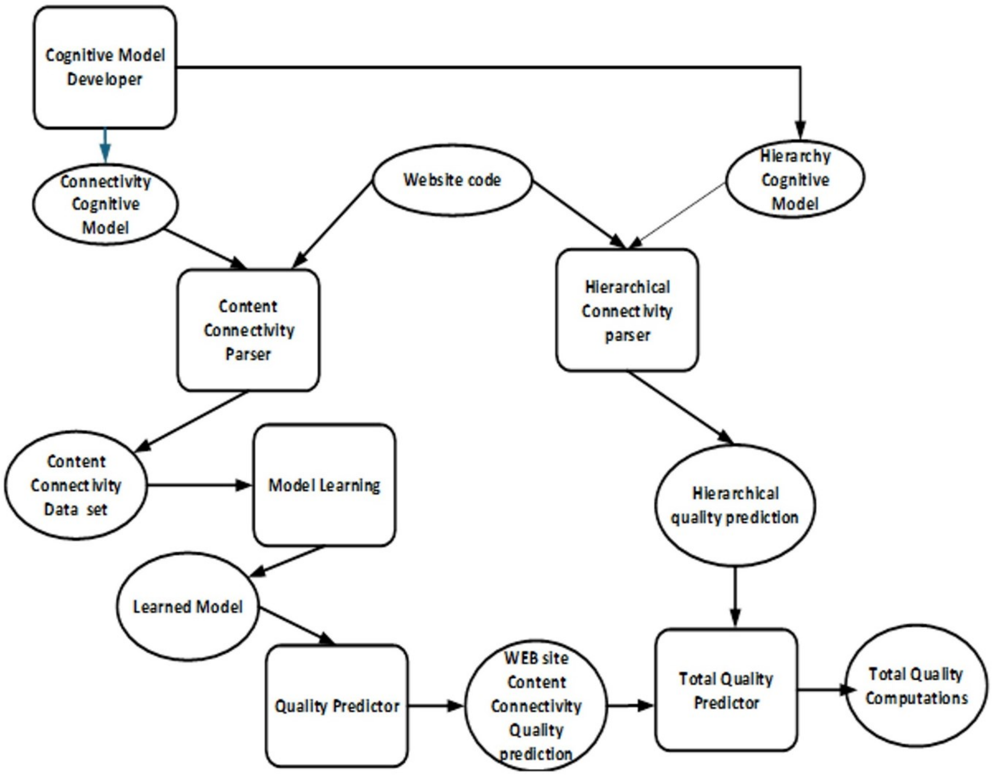

Figure 1. Computing quality of content–methodology.

The quality of content is recognized as encompassing both content connectivity and hierarchical connectivity. The cognitive model used to compute the quality of content connectivity and hierarchical connectivity has been captured and maintained based on expert cognition and the perception of website users.

Two parsers have been used in this model. The first parser scans the website code, generates text, data, and text-data relationships, and computes the percentage of connectivity between all relationships. The related cognitive models are used to compute the quality of all types of connectivity, considering % of connectivity.

One hundred websites have been scanned, and an example set has been created. Using the example set, an optimized neural network is learned by varying the batch size, number of EPOCHS, and data splits. Learning is also done by changing the number of hidden layers, activation functions, and optimizers. The model predicts the quality of the content of any of the new websites.

The second parser scans through the website code and generates hierarchical structures connecting the content designed into the website. The percentage of hierarchical structures is computed based on the standard hierarchical structures. The quality of hierarchical structures is computed using a cognitive model defined by experts and generated according to the users' perceptions.

The total quality of the content is computed as an average of the quality of content connectivity and content hierarchical connectivity.

4. Preparing the example set for computing the quality of connectivity

One hundred websites have been selected, and a parser is run to generate repositories relating to the quality of content hosted on a website. The total percentage of connectivity considering text, data, and text-data connectivity is computed. **Table 1** shows a sample of the example set created.

Table 1. Sample example data set.

Website ID	% Connected Text Counts	% Connected Data Counts	% Connected Text-Data Counts	Average Connectivity	Quality Assessment
1	91	87	77	85	High
2	90	85	80	87	High
3	65	67	69	67	Average
4	67	58	79	68	Good
5	78	88	99	88	High
6	62	78	87	76	Good
7	20	30	40	30	Poor
8	65	80	69	71	Average
9	87	69	70	75	Good
10	46	70	33	50	Poor

The quality assessment is done using the cognitive model shown in **Table 2**. The average connectivity of all three factors is calculated, and this value is used to determine the grade of the content. The cognitive model is developed based on expert perception or users' opinions, which can be changed according to user perception.

Table 2. Cognitive model for computing the quality of connectivity and structural connectivity of the content.

Quality Grading	Poor	Average	Good	High
Percentage Connectivity	< 65	65–75	75–85	85–100
Quality Grading based on %Connectivity	0.00	0.25–0.50	0.50–0.75	0.75–1.00

The cognitive model can be represented empirically using Equation (1) to compute the quality of connectivity and Equation (2) to compute the quality of hierarchical connectivity.

$$\mathrm{QC} = \big((\text{Average Percentage Connectivity} - 64) \times 0.025\big) + 0.25 \tag{1}$$

QC = Quality of connectivity.

$$\mathrm{QHC} = \big((\text{Average Hierarchical Percentage Connectivity} - 64) \times 0.025\big) + 0.25 \tag{2}$$

QHC = Quality of hierarchical connectivity.

5. Investigations and findings

5.1. Computing the connectivity of content

Content is hosted on the WEB regarding text, data, images, videos, 2D/3D graphics, graphs, etc. The quality of the content in basic form can be computed in terms of the hierarchical representation of the content and the connectedness of the content elements. The quality of images, videos, and 2D/3D graphics falls under the quality factor “Multimedia”.

Connections should be made between content elements exposed to users. Text and data must be connected and read appropriately. The computation of the depth of connection includes text, data, and connectors.

Since content and data are dispersed across numerous pages, computing connectivity and building linkages between them is difficult. Connectedness can be developed on a single page, but the total connectedness is the aggregate of all pages. Equation (3) shows that the average content connectedness is the average connectedness of text, data, and text data.

$$\mathrm{AVGCON} = \frac{\%\mathrm{TEXTCON} + \%\mathrm{DATACON} + \%\mathrm{DATA} - \mathrm{TEXTCON}}{3} \tag{3}$$

%TEXTCON is computed by considering all text connectivity and non-connectivity. **Table 3** shows an example of text connectivity.

Table 3. Example text connectivity.

Serial Number	Computing parameter	Count
1.	“Text count”	23,003
2.	“The number of connected text counts”	21,001
3.	“% of Connected Text Counts”	0.91

%DATACON is computed by considering all data connectivity and non-connectivity. **Table 4** shows an example of data connectivity.

Table 4. Data connectivity.

Serial Number	Computing parameter	Count
1	"Data count"	2409
2	"Number of connected data counts"	2109
3	"% of Connected Data Counts"	0.87

%TEXT-DATA is computed by considering all text-data connectivity and non-connectivity. **Table 5** shows an example of data-text connectivity.

Table 5. Text-data connectivity.

Serial Number	Computing parameter	Count
1	"Text-data count"	25,412
2	"Number of connected text-data counts"	18,099
3	"% of Connected Text-data Counts"	0.71

The average connectivity is computed as shown in Equation (4).

$$\mathrm{AVGCON} = \frac{(0.91 + 0.87 + 0.71)}{3} = 0.83 \tag{4}$$

The quality of the content's connectedness is computed using AVGCON and the cognitive table shown in **Table 2**.

Using Equation (1), the quality of the content's factor connectedness is computed as [(83 − 64) × 0.025 + 0.25 = 0.725], which is categorized as "Good".

The parser scans through all the website code, tracing the text elements, data elements, and the connectivity between the text and data, and produces counts and a repository.

The parsers are based solely on static content, which may change dynamically when dynamic content is added to the website. However, the nature of the text and the data will be the same as the content belongs to the same domain.

5.2. Algorithm for computing the connectivity of the content

Algorithm 1, related to the parser, aims to scan all website resource files and generate three repositories related to text, data, and text-data relationships. The repositories are further processed to determine the percentage connectivity, which is then used to calculate the average connectivity. The average connectivity is used to find the quality grade of connectivity using the expert model shown in **Table 2**.

Algorithm 1 Content connectivity parser algorithm

```
Input: Number of resource files
Output: content quality
Process
For each resource file
{
    Locate the Text and Text connectivity and update the Text Connectivity Repository table.
    Find the Data and data connectivity and update the Data Connectivity Repository Table.
    Find the Data and Text connectivity and update the Data and Text Connectivity Repository Table.
}
```

Algorithm 1 (*Continued*)

Generate Connectivity Percentages considering all the repositories
Find the average Connectivity Percentage
Compute content quality by looking into the Content connectivity. Cognitive Table

5.3. Learning a neural network for predicting the quality of connectedness of the content

An example set is created by processing the code of 100 websites using a specially designed parser. Three counts have been computed for each website, considering the sub-factors text, data, and text-data. The average of all three counts is calculated, and the quality is assessed using the cognitive model shown in **Table 2**. A sample of the example set created is shown in **Table 1**.

A neural network is tried with several hidden layers and one input and output layer. The input layer is fed with the percentages of all three connectednesses, and the output layer deals with all four quality grades as described in the cognitive model. The design dealing with one hidden layer yielded the highest accuracy.

The ANN model is trained using 80% of the example set and tested using 20% of the records. No specific bios have been used. The bios are originally taken as 0.0 and later are corrected. The cross-validation method is used for computing accuracy. The loss is computed using the cross-entropy method. The experiment is conducted by changing the number of epochs starting from 100 until the accuracy is improved, keeping the batch size fixed. Several iterations have been carried out, varying the batch size from 2 to 10, and the optimum accuracy achieved has been noted. L2 regularization is included in every layer, considering the overfitting issue. The experiment is conducted by varying the hidden layers within the 1–3 range. Optimum accuracy has been achieved when only one hidden range has been used.

The OS environment, especially the seed for the dataset, is kept constant to obtain the same accuracy for the same trial conducted across multiple trials. Different activation methods have been employed in various layers (ReLU in the input layer, sigmoid in the hidden layer, and SoftMax in the output layer) to address the issue of data nonlinearity. The data has been considered in five splits, and the example records are shuffled in every iteration.

Several model parameters have been explored to achieve the optimal accuracy of the model. The details of the parameters used are shown in **Table 6**.

Table 6. Model parameters.

Layer	**Inputs**	**Outputs**	**Initializer**	**Activation Function**
Input Layer	3	5	Normal	Relu
Hidden Layer	5	4	Normal	Sigmoid
Output Laer	-	4	Normal	Softmax
Batch Size	5	Number of Epochs	600	
Optimiser	Admas	Accuracy	Cross Validation	
Overfitting	L2 Regularisation	Loss function	Categorical Cross Entropy	

The model considers three inputs, five hidden nodes, and four output nodes. The SoftMax activation function in the output layer determines the grade of the content hosted on the website. The model is used to predict the quality of the website.

5.4. Computing the quality of hierarchical connectedness of the text.

A parser is developed that scans through the WEB site code and generates all the hierarchical structures in the text and text data.

13 elements (text, data, and text-data) have been determined to be optimum to be contained in a hierarchy as per the experts in the field.

The parser generates the hierarchies contained in the websites' code, compares them with the standard number of hierarchies, and determines the average percentage of hierarchies, which is used to compute the quality grade of the hierarchies. The hierarchy computation of text, data, data-text, and the average of the same is shown in Equations (5)–(8).

$$\%\mathrm{TEXTHIER} = \frac{\mathrm{TOTTEXTHIER}}{\frac{\mathrm{TOTTEXTELEMENTS}}{13}} \tag{5}$$

$$\%\mathrm{DATAHIER} = \frac{\mathrm{TOTDATAHIER}}{\frac{\mathrm{TOTDATAELEMENTS}}{13}} \tag{6}$$

$$\%\mathrm{DATA} - \mathrm{TEXTHIER} = \frac{\mathrm{TOTDATA} - \mathrm{TEXTHIER}}{\frac{\mathrm{TOTDATA} - \mathrm{TEXTELEMENTS}}{13}} \tag{7}$$

$$\mathrm{AVGHIER} = \frac{(\%\mathrm{TEXTHIER} + \%\mathrm{DATAHIER} + \%\mathrm{DATA} - \mathrm{TEXTHIER})}{3} \tag{8}$$

5.5. Algorithm for computing the quality of content hierarchies

Algorithm 2, related to the parser, aims to scan all website resource files and generate three sets of hierarchies related to text, data, and text-data relationships. The parser counts the number of hierarchies of each type and the percentage of each type by considering the standard hierarchies of each type.

Algorithm 2 Computing the quality of content hierarchies

```
Input: Number of resource files
Output: Number of Hierarchies, Number of expected Hierarchies, Hierarchical content quality
Process
For each resource file
{
   Find the hierarchies considering Data- data-text connectivity and Text connectivity.
   For each of the Hierarchy generated
   {
      Add to the hierarchy repository table if not already present
      Update the Hierarchy if already present
   }
}
Count the number of Hierarchies.
Compute Expected Hierarchies
Compute Quality of Hierarchies
```

5.6. Computing the total quality of content hosted on the websites

The total quality of the content hosted on a website is computed by taking the average quality of connectedness of the content + the quality of connected hierarchies. Algorithm 3 is used for computing this. This is because both sub-factors have equal weight. However, the weights may vary from website to website, in which case the linear regression model involving two variables is to be learned, and the model can be used to compute the overall quality of the website.

Algorithm 3 Total content quality computing algorithm

```
Input: Content Connectivity Quality, Hierarchy Connectivity Quality
Output: Total content quality
Process

If Equal weightage
(
    Total-Content quality = Content Connectivity Quality + Hierarchy connectivity Quality
    Return Total Content Quality
}
```

If the two variables do not have equal weightage, learn a linear regression model involving two variables: text-data connectedness and text-data hierarchical connectedness.

5.7. Linear regression model involving text-data connectedness and text-data hierarchical connectedness

The overall quality of the content

$$OQC = \alpha * \mathrm{QC} + \beta * \mathrm{QHC} + \gamma \tag{9}$$

where QC = quality of text connectedness and QHC = quality of hierarchical text connectedness

The linear regression model using the above data is learned, as shown in Equation (9), which can be used to compute the overall quality of the connectivity. The model is trained using a synthetic dataset sample, as shown in **Table 7**. The values of the coefficients are calculated as

Table 7. Computing the overall quality of the content.

Quality of Text Data Connectedness	Quality Text Data Hierarchical connectedness	Overall quality of the Content	Grade of the Overall quality of the content
0.40	0.75	0.60	Good
0.00	0.25	0.10	Poor
0.25	0.50	0.40	Average
0.75	0.75	0.75	Good
0.80	0.85	0.82	High
1.00	1.00	1.00	High
0.85	0.85	0.80	Hogh
0.25	0.25	0.25	Poor
0.50	0.50	0.50	Average

$\boldsymbol{\alpha}$ = 0.38435627, $\boldsymbol{\beta}$ = 0.54557522, γ = 0.03568 and Mean Square Error (MSE) = 0.0022

With the MSE being negligible, Equation (10) can be used to calculate the overall quality of the content.

$$(OQC) = 0.38435627 * \text{QC} + 0.54557522 * \text{QHC} + 0.03568 \tag{10}$$

6. Results and discussions

6.1. Platform for implementing the model

The Microsoft 11 notebook interface utilizes KERAS and TensorFlow. The software runs on a Dell laptop with a 12th-generation processor and eight cores, as well as two Nvidia GPUs.

6.2. Inputting the data into the model

After every epoch, the example set is shuffled into ten batches of 10 examples. Weights and bios are updated in batches of 5 from each split.

6.3. Generating the text, data and data-text repositories

Parser-1 generates text, data, and text-data relationships in **Tables 8–10**; the parser counts text and data tokens in web resource files. The tokens link text and data and text to data.

Table 8. Text relationships.

Serial number	From text word/String	To Text Word/String
1.	"Registration Number"	"Student Name"
2.	"Registration Number"	"Student Age"
3.	"Registration Number"	"Student Gender"
4.	"Registration Number"	"Department name"
5.	"Registration Number"	"Program name"
6.	"Registration Number"	"Program Year"

Table 9. Data relations.

Serial number	From Data	To Data
1.	"Registration Number"	"SName"
2.	"Registration Number"	"Sage"
3.	"Registration Number"	"PNumber"
4.	"Registration Number"	"SGender"
5.	"Registration Number"	"Dname"
6.	"Registration Number"	"Pname"
7.	"Registration Number"	"Pyear"

Table 10. Data-text relationships.

Serial number	Data	Text
1.	"Regst"	"Registration Number"
2.	"SName"	"Student Name"
3.	"Sage"	"Student Age"
4.	"PNumber"	"Phone Number"
5.	"SGender"	"Student Gender"
6.	"Dname"	"Department name"
7.	"Pname"	"Program name"
8.	"PYear"	"Program Year"

6.4. Generating the overall content connectedness

Using these three tables, the connectedness of content is computed as shown in **Table 11**.

Table 11. Connectedness computations.

Serial Number	Computing parameter	Count
1	"Text count"	23,003
2	"The number of connected text counts"	21,001
3	"% of Connected Text Counts"	0.91
4	"Data count"	2409
5	"Number of connected data counts"	2109
6	"% of Connected Data Counts"	0.87
7	"Text-data count"	25,412
8	"Number of connected text-data counts"	18,099
9	"% of Connected Text-data Counts"	0.71
Average Connectedness	(0.91 + 0.87 + 0.71)/3	0.83

6.5. NN model learning

Tables 12 and **13** illustrate the weights learned for optimal accuracy, with bios set at zero for each output. For 100 to 1000 epochs and 2 to 10 batch sizes, batch sizes of 600 and 5 yielded the best results.

The model is tested with a prediction input of [80, 90, 87], and the quality of content connectedness is computed as "High" with a model accuracy of 98.8% with a standard deviation of 0.11.

Table 12. Weights between inputs and the hidden layer (3 inputs, 5 hidden outputs).

Input Node	Hidden Node 1	Hidden Node 2	Hidden Node 3	Hidden Node 4	Hidden Node 5
Weights between Inputs and the Hidden Layer (3 Inputs, 5 Hidden Outputs)					
Input Node 1	0.04195102	0.02431638	−0.02572782	−0.11830179	0.03076882
Input Node 2	0.04520375	0.02467527	−0.04711938	0.02405597	0.02259666]
Input Node 3	0.03894471	0.01742869	0.01658031	−0.01599376	0.01823624

Table 13. Weights between inputs and the hidden layer (3 inputs, 4 outputs).

Weights between Hidden and Output Layer (5 Hidden and 4 Outputs)				
Hidden Node 1	2.4564118	0.06471535	0.57435024	0.07874147
Hidden Node 2	1.2565916	0.6478258	2.0039039	0.6626415
Hidden Node 3	−0.03126728	0.08539722	0.03843815	0.04503908
Hidden Node 4	−0.03697554	−0.04788164	0.06458279	−0.09061076
Hidden Node 5	1.2820297	1.3023995	3.1787398	1.3352623
Output	−2.121352	−4.2036324	−3.9582913	−3.9582913

6.6. Generating the overall hierarchical content connectedness

The hierarchical depth used for presenting the content is one of the most important issues, as users reading the content hierarchy often need help with collaboration. The deeper the content, the more complex it is to understand.

The 2nd parser scans all resource files and generates the existing text, data, and text-data hierarchies.

Figure 2 shows a sample hierarchy generated from the source files. **Table 6** shows some of the possible hierarchies. From **Table 14**, the average hierarchical connectedness is 91.66. The quality of hierarchical connectedness is rated as HIGH, as indicated by the expert system shown in **Table 2**. Five nodes per hierarchical level have been considered the standard number of nodes, which reflects high-quality structures.

Figure 2. Sample hierarchical diagram.

Table 14. Sample content hierarchies.

Hierarchy Number	Type of Hierarchy	Number of Levels	Number of nodes	Standard Nodes	Percentage of Hierarchical connectedness
1	Data	3	15	15	100.00
2	Data	5	23	25	92.00
3	Text	2	9	10	90.00
4	Text	4	16	20	80.00
5	Data-Text	5	24	25	96.00
6	Data-Text	5	23	25	92.00
Average					91.66

6.7. Accuracy comparisons

MLP model accuracy averages 91.8% with a standard deviation of 0.11%. Jayanthi et al. [28] compared their Extreme Learning Machine (ELM) technique to SVM [29] and Ivory [30]. **Table 15** compares all techniques.

Table 15. Comparison of accuracy of methods used to compute the quality of the websites.

“Method”	“Reference”	“Average Accuracy”	“Standard Deviation”
SVM	[28]	78.00%	0.3%
Ivory	[29]	75.00%	0.7%
ELM	[30]	89.00%	1.2%
NN Model	This paper	91.80%	0.11%

The model was further tested using websites of varying sizes and complexities. The quality of these websites was assessed and compared to the computed quality using the cognitive model and the predicted quality using the NN model. The details are provided in **Table 16**. The computed and predicted quality is seen to be 100% accurate, considering the selected websites.

Table 16. Cognitive model assessment vs. NN-based prediction of the quality of the selected websites.

Web site Number	Website Domian	Number of Pages	Cognitive Model-Based Quality	NN-based predicted Quality
1	Hotel	62	Excellent	Excellent
2	Education	72	Good	Good
3	School	31	Excellent	Excellent
4	Hospital	101	Average	Average
5	E-commerce	138	Very Good	Good

7. Applicability of transfer learning

The proposed paper is related to computing the quality of content hosted on the website, considering the connectedness of text, data, and text-data, as well as the hierarchical connectedness of text and data.

Every website is presented as a set of code elements that contain text and data. For the current research, the input consists of website-related code, which is evaluated based on the quality of its content.

Many models have been available that focus on assessing the quality of websites from the perspectives of text connectedness (Coherence-BERT), data connectedness (GNN), text-data fusion, and the hierarchical structure of text and data (LSTM). The following are the details of the models, which are fully learned based on large-scale synthetic data, though not directly related to website data.

1) Text connectedness using BERT for coherence scoring;
2) Data connectedness with graph neural networks;
3) Text-data fusion for linking structured & unstructured data;
4) Hierarchical structuring using an LSTM-based approach.

These models can be retrained to assess the quality of the websites by generating synthetic data through parsing the website code. The score output by these models needs to be ensembled, and the overall quality outputted through a cognitive model added to the pipeline. This can be done as future research. **Figure 3** explains the concept.

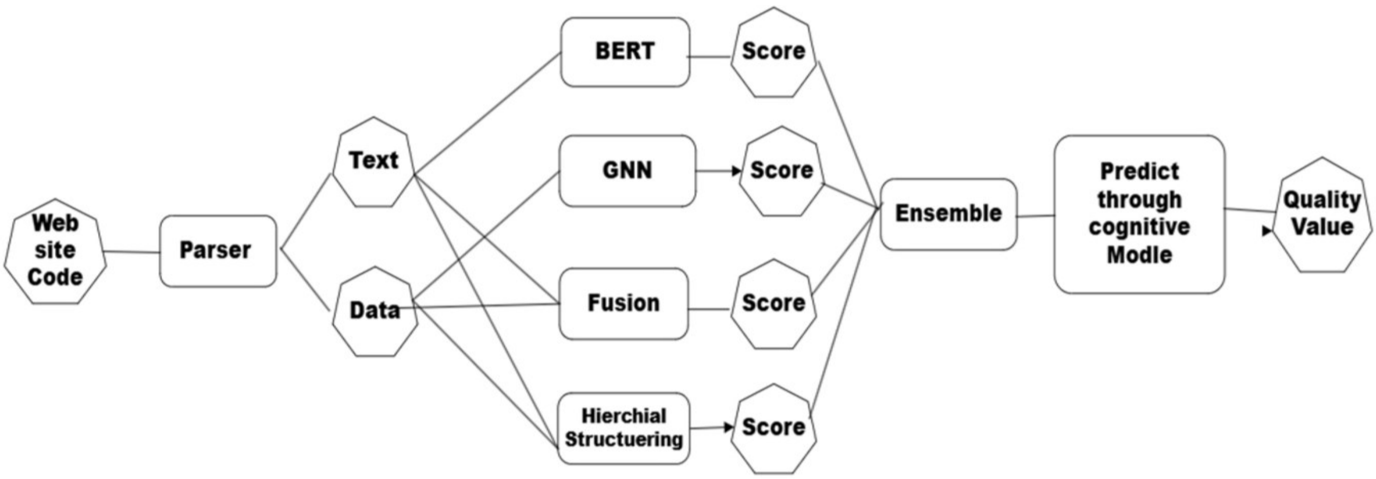

Figure 3. Predicting the quality of websites through transfer learning.

8. Conclusions

It is essential to accurately predict the quality of the content hosted on the website. If the website is not well-designed and developed, it is rarely used. The content hosted on the website has two dimensions: Connectedness and hierarchical structuring. Parsers compute the individual elements of both dimensions by scanning through the website code, and expert systems help compute the quality grade of each dimension. Considering the quality of each dimension, the average quality reflects the overall quality of the entire website. Building an AI model will help predict the quality of a website quite quickly, and such a method is always more reliable. The method proposed in this paper yielded 98.8% with a standard deviation of 0.11.

Author contributions: Conceptualization, KRSJ, STS, HBD, KS, SBJ and BKD; methodology, KRSJ, TM, HBD, and KS; software, STS, HBD, KS, SBJ and BKD; validation, SBJ, and BKD; formal analysis, KRSJ, SBJ, and BKD; investigation,

KRSJ, STS, HBD, and KS; resources, KRSJ, STS, HBD, and KS; data curation, BKD; writing—original draft preparation, KRSJ and SBJ; writing—KRSJ, STS, HBD, KS, and SBJ; visualization, BKD and SBJ; supervision, KRSJ; project administration, KRSJ; funding acquisition, KRSJ. All authors have read and agreed to the published version of the manuscript.

Institutional review board statement: Not applicable.

Informed consent statement: Not applicable.

Conflict of interest: The authors declare no conflict of interest.

References

1. Priya BV, Sastry JKR. Assessment of Website Quality based on Appearance. International Journal of Emerging Trends in Engineering Research. 2019; 7(10): 360–375. doi: 10.30534/ijeter/2019/017102019
2. Priya BV, Sastry JKR. Computing Quality of Structure of a Web-Site. International Journal of Advanced Trends in Computer Science and Engineering. 2019; 8(5): 2142–2148. doi: 10.30534/ijatcse/2019/44852019
3. Priya BV, Sastry JKR. Computing Quality of Navigation Designed into a WEB-Site. International Journal of Emerging Trends in Engineering Research. 2019; 7(11): 466–472. doi: 10.30534/ijeter/2019/117112019
4. Sastry JKR, Sreenidhi N, Sasidhar K. Quantifying quality of WEB site based on usability. International Journal of Engineering & Technology. 2018; 7(2.7): 320. doi: 10.14419/ijet.v7i2.7.10606
5. Priya BV, Sastry JKR. Framework for Assessing the Quality of Multimedia objects Hosted on a WEB-Site. International Journal of Emerging Trends in Engineering Research. 2019; 7(11): 498–505. doi: 10.30534/ijeter/2019/167112019
6. Priya BV, Sastry JKR. Building an Expert System through Machine Learning for Predicting the Quality of a Website Based on Its Completion. Computers. 2023; 12(9): 181. doi: 10.3390/computers12090181
7. Ghalut T, Larijani H, Shahrabi A. Content-Based Video Quality Prediction Using Random Neural Networks for Video Streaming over LTE Networks. In: Proceedings of the 2015 IEEE International Conference on Computer and Information Technology; 2015.
8. Piskozub L. The culture of textual content of a website from the point of view of a specialist editor (Ukrainian). Bulletin of the Book Chamber; 2023.
9. Abiantoro D, Kusumo DS. Analysis of Web Content Quality Information on the Koseeker Website Using the Web Content Audit Method and ParseHub Tools. In: Proceedings of the 2020 8th International Conference on Information and Communication Technology (ICoICT); 2020.
10. Ricca F, Tonella P. Analysis and Testing of Web Applications. In: Proceedings of the International Conference on Software Engineering; 2001.
11. Alwahaishi S, Snášel V. Assessing the LCC Websites Quality. In: Proceedings of the 9th RoEduNet IEEE International Conference; 2010.
12. Hasan L, Abuelrub E. Assessing the quality of web sites. Applied Computing and Informatics. 2011; 9(1): 11–29. doi: 10.1016/j.aci.2009.03.001
13. Singh KK, Kumar P, Mathur J. Implementation of a Model for Websites Quality Evaluation - DU Website. International Journal of Innovations & Advancement in Computer Science. 2014; 3(1).
14. Abduganiev SG. Website Quality Evaluation Methodology Universal Star: 1st point – “Content.” Informatics. 2020; 17(3): 95–112. doi: 10.37661/1816-0301-2020-17-3-95-112
15. Morales-Vargas A, Pedraza-Jimenez R, Codina L. Website quality evaluation: a model for developing comprehensive assessment instruments based on key quality factors. Journal of Documentation. 2023; 79(7): 95–114. doi: 10.1108/jd-11-2022-0246
16. Kardanmoghaddam H, Masoumi M. Website quality assessment. A case study of GSM hosting forum. Facta universitatis-series: Electronics and Energetics. 2022; 35(2): 199–216. doi: 10.2298/fuee2202199k

17. Magalhães LFG, Gonçalves MA, Canuto SD, et al. Quality assessment of collaboratively-created web content with no manual intervention based on soft multi-view generation. Expert Systems with Applications. 2019; 132: 226–238. doi: 10.1016/j.eswa.2019.04.053
18. Raghavarao YV, Sasidhar K, Sastry J, et al. Quantifying quality of WEB sites based on content. International Journal of Engineering & Technology. 2018; 7(2.7): 138. doi: 10.14419/ijet.v7i2.7.10280
19. Muirhead J. Assessing Online Content Quality through User Surveys and Web Analytics. In: Proceedings of the 2019 Conference on Human Information Interaction and Retrieval; 2019.
20. Morales-Vargas A, Pedraza-Jiménez R, Codina L. Website quality: An analysis of scientific production. Informational pluralism. 2020; 29(5). doi: 10.3145/epi.2020.sep.08
21. Rekik R, Kallel I, Casillas J, et al. Assessing web sites quality: A systematic literature review by text and association rules mining. International Journal of Information Management. 2018; 38(1): 201–216. doi: 10.1016/j.ijinfomgt.2017.06.007
22. Morales-Vargas, Alejandro; Pedraza-Jiménez, Rafael; Codina, Lluís (2020). “Website quality: An analysis of scientific production”. Professional de la information, v. 29, n. 5, e290508. https://doi.org/10.3145/epi.2020.sep.08
23. Zhang J, Song Z, Li Y. UGC content quality assessment method and device and electronic device. CN201910535860.8, 20 June 2019.
24. He L, Tang L, Wang N. Quality Assessment Method of Web Documents Based on Random Forest. In: Proceedings of the 2017 5th International Conference on Mechatronics, Materials, Chemistry and Computer Engineering (ICMMCCE 2017); 2017.
25. Mavetera P, Lubbe S, Meyer J. A student perspective on information quality of websites. The African Journal of Information SystemsThe African Journal of Information Systems. 2017; 9(3).
26. Dalip DH, Lima H, Gonçalves MA, et al. Quality assessment of collaborative content with minimal information. IEEE; 2014.
27. Klink J, Bardowski P, Uhl T. Content-Agnostic Web Browsing Quality Assessment. In: Proceedings of the 2018 26th International Conference on Software, Telecommunications and Computer Networks (SoftCOM); 2018.
28. Jayanthi B, Krishnakumari P. An intelligent method to assess webpage quality sing extreme learning machine. International Journal of Computer Science and Network Security. 2016; 16(9): 81–85.
29. Huang GB, Ding X, Zhou H. Optimization method based extreme learning machine for classification. Neurocomputing. 2010; 74(1–3): 155–163. doi: 10.1016/j.neucom.2010.02.019
30. Huang G, Zhou H, Ding X, et al. Extreme Learning Machine for Regression and Multiclass Classification. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics). 2012; 42(2): 513–529. doi: 10.1109/tsmcb.2011.2168604