Centralized Explain Items Semantic Method and Its Application

Jiarui Zhang
Department of Computer Science and Technology, Hefei University, Hefei 230601, Anhui, China
(E-mail: zhangjiaru099@163.com)

Abstract
In order to solve the issues brought by Large-scale data exchange such as the “highly automated process”, “high transmission capacity” and the “high data packet define efficiency”, proposes the centralized explain items semantic and the design of the “items semantic converter (ISC)”. This forces the “information producer” and the “information consumer” to interact on demand through “requirement of exchange” and the “data of exchange”, afterward with the support of the “items conversion” and the “data-elements retrieval”, enables information sharing without attachment of the “meta-data”. Furthermore, the differences in processes between the method aforementioned and the " dispersion explain items semantic " method in the “large-scale data exchange applications” had been analyzed, then conducted a contradistinction experiment in terms of the “transmission capacity” and "data package analysis efficiency”. As the result of the experiment and analysis, the proposed method is possessed with “completed automatic process”, “higher transmission flow (average higher 52.72%)”, “higher packet parsing efficiency (average higher 27.81%)”. This indicates that the proposed method is a newer and better solution of “large-scale data exchange”.

Key words: Explain Items Semantic, Dispersion Explain Items Semantic, Centralized Explain Items Semantic, Data of Exchange, Items Conversion.

1. INTRODUCTION
At present, the whole society informatization level, total information faster growth. However, because of the lack of top-level design and unified planning, the formal use of business system general form “information island”, information sharing across the system more difficult, especially in the “wisdom government”, “intelligent decisions”, “knowledge discovery” and other large-data technology to support the application needs urgent today, the scope and scale of information sharing is more and more big, the significance of technological innovation in information sharing is particularly important.

Data exchange is the major technical measure of information sharing across systems. During the process of data exchange, the information of transmission across systems is mainly a collection data items. Meanwhile, both parties of the exchange must develop a coherence on the understanding of the items’ definition in order to achieve information sharing. Therefore, the issues generated within the Explain Items Semantic are the inevitable obstacles to overcome (Laurel and David, 2012; Arnon and Len, 2014). At present, the technology of the data exchange in general can be divided into: “the data exchange technology based on middleware” (Liu and Chen, 2009; Ming and James, 2013), “the data exchange technology based on ESB (Enterprise Server Bus)” (Li and Yang, 2006; Pan and Wang, 2014; Wang and Xu, 2012; Andreas and Mohammed, 2013), “the data exchange technology based on meta-data” (Mikolaj and Leszek, 2013; Rada and Leonid, 2012; Tan and Zhang, 2013; Kathryn and Joop, 2011). They solve across systems information sharing method is summarized as follows:

- “The data exchange technology based on middleware” and "the data exchange technology based on ESB": They means the manual process of items defining on each related items between the parties of the information producer and consumer. This is the "artificially explain items semantic" method.
- “The data exchange technology based on meta-data”: Which firstly requires the producer and consumer to determine the topic of the exchange, the producer will then carry out the description process of the items’ definition within the transferred information through the meta-data. The consumer would then understand the definition of each item by analyzing the meta-data. This is the "dispersion explain items semantic" method.

As informatization development up to current timeline, the practical application of large-scale data exchange mostly consist the major characteristics of “inability to predict the number of exchange-node in the perimeter of the exchange, inability to identify the parties related to the exchange beforehand, inability to recognize the exact date and time of the exchange and the inability to identify the content of the exchange”. Obviously, since this is a Random exchange which is driven by sharing application and requires a high level of automatic process, high transmission capacity, and high data packet define efficiency, the artificially explain items semantic method cannot be applied to this type of data exchange. On the other hand, the application of dispersion explain items semantic requires attachment of meta-data to all items within the exchange, which
would inevitably cause the exchanging data package to expand significantly, and would lead to the result of occupying a larger bandwidth, low transmission efficiency and low resolve efficiency which conflicts with the basic requirement of large-scale data exchange. Currently, the mainstream of the research are focused on methods to raise the resolve efficiency of data packages, and the industry is still a lack of solutions to solve the above issues.

For that reason, this essay proposes a new method of "centralized explain items semantic". In between "customers" and "producers" of information, through the "exchange request" and "exchange data" on-demand interaction, the formation of "random data exchange" mechanism. By the method provided by the "data item batch conversion" and "data-element batch retrieval", realized without additional "meta-data", do not need to parse the packet can achieve the goal of information sharing. Can solve the problem of data exchange process fully automated, increase efficiency of the flow and parsing. It is a better solution of large-scale data exchange. In this paper, the "centralized explain items semantic method" is detailed design, the application of this method in the large-scale data exchange are introduced, the application process of this method and "dispersion explain items semantic" method are compared, between this method and "dispersion explain items semantic methods" with "flow" and "packet parsing efficiency" has carried on the contrast experiment, the "centralized explain items semantic method" effect is obvious.

2. CENTRALIZED EXPLAIN ITEMS SEMANTIC

2.1. The Semantic Model of Item

The semantics of items can be separate as "value semantics" and "extension semantics"; where "value semantics" can be further divided into 3 parts of the items as Value, Type, and Range. The extension semantic is a type of semantic circulated from items of business systems and context of the data table, which can be described through items label and property. Such examples can be found in table 1.

<table>
<thead>
<tr>
<th>Number</th>
<th>Items name</th>
<th>Items value</th>
<th>Data type</th>
<th>Range</th>
<th>Items label</th>
<th>Data property</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>name</td>
<td>San Zhang</td>
<td>String</td>
<td>Length of not more</td>
<td>Employee name</td>
<td>name</td>
</tr>
<tr>
<td>2</td>
<td>ID</td>
<td>342601197808240611</td>
<td>String</td>
<td>GB/T 11643-1999</td>
<td>Employee ID</td>
<td>number</td>
</tr>
<tr>
<td>3</td>
<td>sex</td>
<td>1</td>
<td>Numberal</td>
<td>GB/T 2261.1-2003</td>
<td>Employee sex</td>
<td>code</td>
</tr>
<tr>
<td>4</td>
<td>startday</td>
<td>2010-07-16</td>
<td>Date</td>
<td>YYYY-MM-DD</td>
<td>Date of employees into the company</td>
<td>date</td>
</tr>
</tbody>
</table>

Item "sex" has a value semantic of <1, numberal,GB/T 2261.1-2003>, and an extension semantic of <employee sex,code>. Through the combination of value and extension semantic the semantic of Item "sex" can be completely described as: Item "sex" refers to the staff's gender, while its value "1" refers to the value: male, which possesses the reference number of 1 in the data dictionary "GB/T 2261.1-2003". Thus, as the result the item semantic model are shown below:

![Figure 1. The model of item semantic](image)

During cross-system data exchange, it is typical to only transfer the items' name and value and ignore the other parts of the items semantic. This causes the confusion and errors when the items are transferred to different system. In order to enable both parties to the exchange to accurately understand the definition of the items, in order words: the achievement of information sharing, whether the method of artificially explain items semantic, dispersion explain items semantic or the Centralized Explain Item Semantic Method, data exchange must attempt to fulfill each item's "type", "range", "label" and "property". The items with completed semantic may be characterized as "data-element".

2.2. Centralized Explain Items Semantic and Process

Centralized explain items semantic refers to the process of establishing a "public implementation" in exchange domain space. This "public implementation" provides all the parts of item semantic except the value part. During cross-system information transmission, the consumer and the producer would combine each value of the exchange items with the support of the "public implementation" to achieve the complete explanation of
each item semantic. This "public implementation" take the "code of data -element" as a key value and enable management, is characterized as "Items Semantic Converter (ISC)".

The process of the Centralized Explain Item Semantic are shown below:

![Diagram](image)

**Figure 2**: The process of centralized explain items semantic

Explanation:

The process of the Centralized Explain Item Semantic are shown below:

①: Producer uses ISC to convert the item name of the pre-exchange data into the code of data-element.

②: Producer delivers the converted pre-exchange data to the consumer.

③: Consumer uses the code of data-element in ISC to execute the retrieval process, achieve the understanding of items.

Apparently the Centralized Explain Item Semantic isn't necessary to apply on every item, instead, it only requires the producer to convert the item names into the code of data-element. This process would not enlarge the size of the data package and would not raise the requirement to the bandwidth, which is a few advantages to support the achievement of High transmission capacity. Furthermore, after the consumer party receives the data, they may initiate the batch retrieval process through the code of data-element in ISC without resolving the metadata.

### 2.3. The Achievement of Centralized Explain Items Semantic

From figure 1, it is obvious that the ISC must define the "type", "range", "label" and "property" of each item within the pre-exchange data. Amongst the definition, "type" and "range" can be described through ordinary data structure, while "label" and "property" are possessed with facts and issues stated below:

- Items label is the modifier-head construction which consists of "attributive adjective" and "headword". The "attributive adjective" serves as a constant to the "headword". Take the item "staff sex" above for example, where "staff" is the attributive adjective and "sex" is the headword. In this case, the term "staff" is the constraint to the term "sex".

- When ISC is treated as a cross-field tools design, it has to adapt to the environment of a lot of attributive adjectives, with the additional object classification and hierarchy-relation among the attributive adjectives.

- Due to the highly variated human languages, "attributive adjectives" may be linked to a high amount of synonyms, while the semantic between these synonyms shares the same importance.

- As the attributive adjectives and the headwords from the items label, it is necessary to develop a relevant relationship between the attributive adjectives and the headwords.

- The categorization issue of the headwords. Since the categorization of headwords is not unlimited, instead it is based on the combination of convention and the value of items.

In order to bring the ISC into practical usage, a few problems must be resolved such as the categorization and hierarchy-relation of the attributive adjectives, the synonyms imputation of the attributive adjectives and finally the relating process of attributive adjectives and headwords.

#### (1) Categorization of attributive adjectives

With the inspiration of how the OOAD master PeterCoad use business model to categorize his objectives (while the business model separates its objectives into four sectors of "people", "location", "event" and "goods"), the attributive adjective is separated into 6 sectors which listed below:

A: representing the categorization of people  
B: representing the categorization of organizations  
C: representing the categorization of geographic locations  
D: representing the categorization of either the objects are tangible or not  
E: representing the categorization of the events/activities  
Q: others
(2) Organization, categorization and coding of the attributive adjective

By constructing the six sectors of attributive adjectives into classification type of semantic field, and categorize the attributive adjective based on their hierarchy relations. Meanwhile, in order to avoid affecting the indexing efficiency, the numbers of layers in the semantic field are recommended to avoid exceeding a certain number.

![Diagram of the classification type semantic field of A]

Figure 3. The example of classification type semantic field of A

The following format should be used to code the leaf node of the semantic field:
- Format 1: One unit of capital English letter - 5 digits number
- Format 2: One unit of capital English letter - 2 digits number - 5 digits number

Note the following rules:
- "One unit of capital English letter" indicates the category of the attributive adjectives, with the value of A, B, C, D, E, Q;
- Secondly "2 digits number" indicates the code of the second layer within any classification type of semantic field, which is reconstructed during the process of reconstruction of the leaf node, with a restraint of 00–99;
- "5 digits number" indicates the codes of the attributive adjectives, which is produced during the process of the examination of the registered items, with a restraint of 00000–99999.

(3) Synonyms indexing of attributive adjective

The treatment of the synonyms are related to the languages in use, for example, if it is an English term then the "WordNet" will be used to directly retrieve the synonyms. In the other case as this part focus in the Chinese language would use the dictionary method instead, which efficiently, accurately and comprehensively index all the synonyms respective to the targeted Chinese terms though the Mandarin synonyms dictionary. Use the 12 categories and 5 layers structure presented in the "HIT Lab Tongyici Cilin (Extended)», and the 77343 Chinese Thesaurus Cilin to construct the synonyms dictionary. The retrieval structure and process of Chinese synonyms can be found in figure 4 and figure 5:
In the classification type of semantic field, synonyms will be collected into the non-differentiated node and possessed with the same node code. Only the synonyms on the leaf node would be collected in this manner.

(4) Headword categorization

Based on the convention combined with practical application, then form the headword categorization and coding while keeping the ability to expand. These process can be seen in the following list:

**Table 2.** The central word property classification and coding table

<table>
<thead>
<tr>
<th>Code</th>
<th>Property classification</th>
<th>Code</th>
<th>Property classification</th>
<th>Code</th>
<th>Property classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>name</td>
<td>01</td>
<td>code</td>
<td>02</td>
<td>Description/explanation</td>
</tr>
<tr>
<td>03</td>
<td>number</td>
<td>04</td>
<td>amount/value</td>
<td>05</td>
<td>date</td>
</tr>
<tr>
<td>06</td>
<td>time</td>
<td>07</td>
<td>unit</td>
<td>10</td>
<td>indicator</td>
</tr>
<tr>
<td>08</td>
<td>percent</td>
<td>09</td>
<td>ratio</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(5) The connection between attributive adjective and the headword

Develop an independent data table and code the headword. Afterward, construct a connection between the leaf node of the semantic field and the headword as below:
(6) Main data structure of the ISC

The main data model of ISC are mainly formed by the data table below:

- Recorded data-elements number, items type and the range table
  `ItemSemanticMoel={
    dataelecID; //vchar(16),Data-elements number, main key,unique
    itemtype; //data type,related to the type of the database
    valuearea  //vchar(50),the restraints of the database(may be a specified dictionary)
  }

The format of the data-elements code is:
Data-elements code = corresponding semantic field leaf node coding-two-digit number- two-digit number;
While the first “two-digit number” represents the serial number of the headword (see figure 6), with a restraint of 00~99;
As the second “two-digit number” represents the categorization serial number of the headword (see table 2), which number is produced automatically during the examination process of against the registered item, with a restraint of 00~99.

- Semantic field data table
- Headword coding table
- Categorization coding table
- Data dictionary related to the restraints
- Synonyms dictionary and categorize table of the attributive adjectives

(7) Function of ISC

The major purpose of ISC is to provide the following function to the “producer”, “consumer” and the “ISC manager”:

ISC mainly provide the service of “registration”, “conversion” and the “retrieval”. While:
- Conversion: Enable the “producer” or the “consumer”: to convert the items name within the pre-exchange data into data-elements number. The “producer” or the “consumer” would search the existing data-elements within the ISC, then select the data-elements which correspond to the requirement to convert the item into data-elements number according to the principle which stated the equal importance amongst the semantic of synonyms.
- Registration: In the case where the ISC are the lack of the required data-elements, the “producer” and the
“consumer” may register items on the ISC. This function allows the producer and the consumer to define the specific data within ISC, and turning the item into data-elements with a unique serial number and completed semantic through the examination of the “manager”. With the “registration” function, the issue of identifying the sources of data-elements may be resolved, while the storage of the data-elements may extend towards perfection in a stabilized manner through operation techniques such as the classification type of semantic field and the collection of the synonyms. Through constant usage, the data-elements within the ISC may gradually fulfill the requirements in utilization as the frequency of requiring the user to registry an item gradually decrease to a stable point.

- Retrieval: Enable the producer and consumer to search the corresponding semantic to the targeted items through the retrieval of data-elements number. The process of retrieval can be found below:

![Figure 8. To retrieve items semantic by Data-Element Number](image)

- The manager of the ISC would examine the registered items, take clustering operation on the leaf node of the semantic field. The items will be then presented to the ISC database after being authorized. The ISC manager may initiate the enable, disable, and forbidding treatment against any data-elements.

In order to achieve convenience and efficiency, ISC supports two conversion/retrieval pattern as “online batch conversion/retrieval” and “interface batch conversion/retrieval”, which increases the efficiency of the conversion/retrieval and making the ISC to possess the ability of utilizing as an independent application of an application platform to settle in the exchange area.

3. The application of Centralized Explain Item Semantic in large-scaled data exchange

3.1. Large-Scale Data Exchange Formulation

Base on the previous statement, large-scale data exchange is sharing-application driven data exchange. It is not able to arrange and customize the specific exchange task prior to the execution of exchange. Hence, the information consumer and producer interactive randomly for data sharing by applying "exchange requirement" and "exchange data". Meanwhile, this interaction will make data sharing be information complete automatic, high transmission capacity and high resolve package speed when using ISC, the formula is shown below:

\[ f_k(EN_i, EN_j) = \langle EO\text{rder}_k, ED\text{ata}_k, ISC \rangle \] ...............................

In this formula, \( f_k \) ( \( EN_i, EN_j \) ) represents the \( k^{th} \) exchange-task between the exchange-node of \( EN_i \) and \( EN_j \) while \( EO\text{rder}_k, ED\text{ata}_k \) indicate the "requirement of exchange" and the "data of exchange" corresponding to \( f_k \) respectively.

Definition 1: “exchange requirement(EOrder)” is representing the specific structure of “within a limited time, retrieving the specifically structured and viable information from certain nodes”.

\[ E\text{Order} = \langle OrderID, EN_j, EN_i, At1, \{EN_j.ItemName\}, EXP(EN_j.ItemName, value) \rangle \] ...............................

Among the factors, The OrderID is the number of EOrder and is globally unique. While \( EN_j, EN_i, \) and \( At1 \) represents the source node, target node and desired time limit respectively. The \( \{EN_j.ItemName\} \) represents
the collection of the local items in the consumers of information resource(ENj). Finally the EXP(ENj, ItemName, value) is the expression integrated from "ENj, ItemName" and the "value".

Definition 2: "Exchange data (EData)" means “the transmission of specific structure and value from target node to source node within the time limit agreed”.

EData =< DataID, OrderID, ENi, ENj, At2, {< ENj, ItemName, value >}> ..................................................(3)

Among the factors, DataID is a number of the EData and is globally unique. The connection through the OrderID the EOrder At2 is feedback time. {< ENj, ItemName, value >} is the key-value pair collection of ENj, ItemName and the value.

![Figure 9. Application chart of the large-scale data exchange](image)

### 3.2. Large-Scale Data Exchange Application Chart

In practical application, the direct data transform between producers to consumers usually change into both direction transform between a Peripheral System and Message Processing Center. The reason for doing this is seeking stand, efficiency and low coupling with operation system.

From this figure, the peripheral system is to have exchange order and exchange data's customize, convert, receipt, and understand, which aim to exchange data with the database. Message Server centre is service exchange order convert, exchange data convert and accept. ISC provides peripheral system registration, convert, and data-elements index service. In addition, ISC provides item semantic to the centre database. Data dictionary provides code mapping service to the peripheral system, ISC system, and centre database. Control monitoring aims to achieve management to the peripheral system. Secondly, the maintenance of characters, users, authorities, and passwords. Thirdly, the management of the message service center and finally the maintenance of the central databases.

![Figure 10. Process model of the large-scale data exchange](image)
3.3. Large-Scale Data Exchange Process Model Based on ISC

Large-scale data exchange model can be developed by EOrder and EDATA: 
From this figure,  
① “consumer” express their requirement by using ISC following data sharing requirement. Then have EOrder

\[
EOrder_k = < OrderID, EN_i, EN_j, At1, \{dataelecID\}, EXP(dataelecID, value) > ....................(4)
\]

② “consumer” send exchange order EOrder_k to “producer”.  
③ “producer” understands exchange order by retrieval ISC.  
④ “producer” get matching data exchange order data from local database.  
⑤ “producer” convert matching data into exchange data EData_k by ISC.

\[
EData_k = < DataID, OrderID, EN_i, EN_j, At2, \{< dataelecID, value > \} > ............................(5)
\]

⑥ “producer” send exchange data EData_k to “consumer”.  
⑦ “producer” understands exchange data EData_k by retrieval ISC.  
⑧ “consumer” write data exchange into local database.

The stage ①③⑤⑦ will automatically complete when using the connect batch conversion/retrieval function of ISC. Which improve the efficiency of conversion/retrieval and simplify the exchange process.

4. APPLICATION PROCESS COMPARISON OF CEIS AND DEIS

Notes:  
DEIS: dispersion explain items semantic, CEIS: centralized explain items semantic.

The following be dispersion explain items semantic, centralized explain items semantic of two different methods are used in large-scale data exchange process made a comparison:

<table>
<thead>
<tr>
<th>Step</th>
<th>Process</th>
<th>Measures</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td>The two sides manual agreed to exchange topic.</td>
<td>manual</td>
<td>Such as: personal tax information</td>
</tr>
<tr>
<td>Step 2</td>
<td>“producer” prepares original data</td>
<td>automatic</td>
<td>Prepare the data based on local items semantics through exchange bridging</td>
</tr>
<tr>
<td>Step 3</td>
<td>“producer” organizes meta-data</td>
<td></td>
<td>Organizes meta-data term by term, while setting up items with serial number to their corresponding data dictionary</td>
</tr>
<tr>
<td>Step 4</td>
<td>Transmit the data package with the attachment of meta-data to the “message process centre”</td>
<td></td>
<td>Transmission of data package</td>
</tr>
<tr>
<td>Step 5</td>
<td>“message process center” sending the data package with meta-data to the “consumer”</td>
<td></td>
<td>Transmission of data package</td>
</tr>
<tr>
<td>Step 6</td>
<td>“Consumer” analyze the meta-data and receives the semantic of the items</td>
<td></td>
<td>Receive the items semantic by analyzing the meta-data</td>
</tr>
</tbody>
</table>

When comparing “table 3” to “table 4”: “Step 1” is corresponding to “step a to step e”, while the former exchange the theme manually, the later exchange the data through “exchange order” automatically and accurately. Apparently the later are possessed of the advantage in completely automatic. “Step 6” and “step 7” reflected two different understandings to the transferring information. Since the later method is based on the “connected batch retrieval service” from ISC to operate automatically, it shows a rather higher efficiency in analyzing comparing to the former method using the “meta-data”. “Step 1” with “step 4” and “step 5” with “step 6” reflected on two different methods’ transmission capacity, which will be analyzed and compared through the following experiment.
Table 4. The application process of centralized explain items semantic method

<table>
<thead>
<tr>
<th>Step</th>
<th>Process</th>
<th>Measures</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>&quot;consumer&quot; organizes &quot;exchange order&quot; based on local items semantic</td>
<td>automatic</td>
<td>By &quot;Eorder&quot; accurately it determines the exchange content</td>
</tr>
<tr>
<td>2</td>
<td>&quot;consumer&quot; converts &quot;exchange order&quot; by ISC</td>
<td></td>
<td>Prepare the data based on local items semantics through exchange bridging</td>
</tr>
<tr>
<td>3</td>
<td>&quot;producer&quot; sends &quot;Eorder&quot; to &quot;massage process centre&quot;</td>
<td></td>
<td>Converting the items name to data-elements code through the connected batch conversion function of ISC</td>
</tr>
<tr>
<td>4</td>
<td>&quot;massage process centre&quot; sends &quot;Eorder&quot; to &quot;producer&quot;</td>
<td></td>
<td>Transmission of data package</td>
</tr>
<tr>
<td>5</td>
<td>&quot;producer&quot; understands &quot;Eorder&quot; by using ISC</td>
<td></td>
<td>Transmission of data package</td>
</tr>
<tr>
<td>6</td>
<td>&quot;producer&quot; prepare original data</td>
<td>automatic</td>
<td>Achieve the understanding of all items semantic through the connected batchretrieval function of ISC</td>
</tr>
</tbody>
</table>

5. EXPERIMENTAL ANALYSIS

Experiments topic: Contradistinction experiment in transmission flow and the data package analysis efficiency of the dispersion explain items semantic method and the centralized explain items semantic method.

Experiments environment:
(1) Front Node:
Amount: two, representing the consumer and the producer of the data
Environment:
OS: Windows XP/32bit
Local memory: 2G
CPU : E3200
DBMS: MySQL 5.5
Browser: 360 Browser.
(2) Message Processing Center:
Amount: one
Environment:
OS: Windows XP/32bit
Local Memory: 2G
CPU : E3200
DBMS: MySQL 5.5
(3) Network and Bandwidth:
Network: LAN
(4) Data transmission measure: WebService
Experiment instruction: The items of meta-data form based on the《GB/T 21063.3-2007》standard.
Experiment data: See the table below.

Table 5. Contradistinction experiment in TC and the DPAE between the DEIS and the CEIS method

<table>
<thead>
<tr>
<th>Number</th>
<th>IndexA</th>
<th>IndexB</th>
<th>IndexC</th>
<th>IndexD</th>
<th>IndexE</th>
<th>IndexF</th>
<th>IndexG</th>
<th>IndexH</th>
<th>IndexI</th>
<th>IndexJ</th>
<th>IndexK</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>632832</td>
<td>1557504</td>
<td>35975</td>
<td>53663</td>
<td>7059.86</td>
<td>127022</td>
<td>1026048</td>
<td>23700</td>
<td>35352</td>
<td>10716.52</td>
<td>101947</td>
</tr>
<tr>
<td>2</td>
<td>1265880</td>
<td>3115458</td>
<td>83908</td>
<td>105335</td>
<td>6601.23</td>
<td>379800</td>
<td>2051868</td>
<td>55263</td>
<td>69375</td>
<td>10156.45</td>
<td>201056</td>
</tr>
<tr>
<td>3</td>
<td>1880361</td>
<td>4675512</td>
<td>124875</td>
<td>159975</td>
<td>6601.23</td>
<td>379800</td>
<td>2051868</td>
<td>55263</td>
<td>69375</td>
<td>10156.45</td>
<td>201056</td>
</tr>
<tr>
<td>4</td>
<td>2531328</td>
<td>6230016</td>
<td>187258</td>
<td>208094</td>
<td>6402.72</td>
<td>379800</td>
<td>2051868</td>
<td>55263</td>
<td>69375</td>
<td>10156.45</td>
<td>201056</td>
</tr>
<tr>
<td>5</td>
<td>3164160</td>
<td>7785512</td>
<td>221070</td>
<td>255550</td>
<td>6629.94</td>
<td>379800</td>
<td>2051868</td>
<td>55263</td>
<td>69375</td>
<td>10156.45</td>
<td>201056</td>
</tr>
<tr>
<td>6</td>
<td>3796992</td>
<td>9345024</td>
<td>270750</td>
<td>321161</td>
<td>6414.80</td>
<td>379800</td>
<td>2051868</td>
<td>55263</td>
<td>69375</td>
<td>10156.45</td>
<td>201056</td>
</tr>
<tr>
<td>7</td>
<td>4429824</td>
<td>10002128</td>
<td>328674</td>
<td>366618</td>
<td>6371.17</td>
<td>886200</td>
<td>7182636</td>
<td>216523</td>
<td>241530</td>
<td>9670.99</td>
<td>692347</td>
</tr>
<tr>
<td>8</td>
<td>5062656</td>
<td>12460032</td>
<td>357370</td>
<td>434442</td>
<td>6393.76</td>
<td>1012800</td>
<td>8208684</td>
<td>235436</td>
<td>286211</td>
<td>9705.14</td>
<td>790747</td>
</tr>
<tr>
<td>9</td>
<td>5695788</td>
<td>14017536</td>
<td>413560</td>
<td>503286</td>
<td>6212.37</td>
<td>1139400</td>
<td>9234732</td>
<td>272453</td>
<td>331562</td>
<td>9429.88</td>
<td>889147</td>
</tr>
<tr>
<td>10</td>
<td>6336320</td>
<td>15575040</td>
<td>467957</td>
<td>559813</td>
<td>6165.11</td>
<td>1266000</td>
<td>1026840</td>
<td>308520</td>
<td>369080</td>
<td>9351.12</td>
<td>987547</td>
</tr>
<tr>
<td>Average</td>
<td>3479614</td>
<td>8566617</td>
<td>249203</td>
<td>296794</td>
<td>6494.02</td>
<td>696342</td>
<td>5644214</td>
<td>162675</td>
<td>195547</td>
<td>9917.52</td>
<td>544818</td>
</tr>
</tbody>
</table>

Notes:
In the experiments, TC: transmission capacity, DPAE: data package analysis efficiency, DEIS: dispersion explain items semantic, CEIS: centralized explain items semantic.
In the experiment data, number of records = 300, 600, 900, ..., 3000, each record number of items = 6, no additional dictionary.
“IndexA”: The exchange of transmission data entity bytes, unit byte;
“IndexB”: The number of bytes of data packets dispersion explain items semantic method, unit byte;
“IndexC”: The transmission time between the consumer front system and message processing center dispersion explain items semantic method, unit ms;
“IndexD”: The transmission time between the message processing center and producer front system dispersion explain items semantic method, unit ms;
“IndexE”: The transfer traffic dispersion explain items semantic method, unit bytes/sec;
“IndexF”: The packet parsing time dispersion explain items semantic method, unit ms;
“IndexG”: The number of bytes of data packets centralized explain items semantic method, unit byte;
“IndexH”: The transmission time between the consumer front system and message processing center centralized explain items semantic method, unit ms;
“IndexI”: The transmission time between the message processing center and producer front system centralized explain items semantic method, unit ms;
“IndexJ”: The transfer traffic centralized explain items semantic method, unit bytes/sec;
“IndexK”: The packet parsing time centralized explain items semantic method, unit ms.

\[
\text{Index}E_i = \frac{\text{Index}A_i}{(\text{Index}C_i + \text{Index}D_i)} \times 1000; \quad \text{Index}J_i = \frac{\text{Index}A_i}{(\text{Index}H_i + \text{Index}I_i)} \times 1000; i \text{ is the row number, } i = 1, 2, ..., 10.
\]

Experiment results:

![Figure 11. Transmission flow comparison](image-url)
Figure 12. Data package parsing time comparison

Transmission flow ratio = \frac{\frac{\sum_{i=1}^{10} \text{Index}_i}{10}}{\frac{\sum_{i=1}^{10} \text{Index}_E}{10}} \times 100\% = \left( \frac{9917.52 - 6494.02}{6494.02} \right) \times 100\% = 52.72\% ;

Parsing efficiency ratio = \frac{\frac{\sum_{i=1}^{10} \text{Index}_F}{10}}{\frac{\sum_{i=1}^{10} \text{Index}_K}{10}} \times 100\% = \left( \frac{696342 - 544818}{544818} \right) \times 100\% = 27.81\% .

6. CONCLUSIONS

This essay presented a centralized explain item semantic method, and researched on the key techniques which achieve the method. A contradistinction experiment and analysis has been conducted on the comparison of the aforementioned method and the existing explain-item-semantic by a meta-data method in a large-scale data exchange situation. As a result of the contradistinction analysis, the centralized explain items semantic method are possessed of a higher automatic level. And as a result of the experiment, the centralized explain items semantic operates with a relatively higher transmission capacity of 52.72% and higher data package analysis efficiency of 27.81% in average. Furthermore, the aforementioned method has been practically applied to more than ten projects in practice, which indicates the ISC are able to “self-develop” through registration of items. Gradually, as the ISC provides services of items conversion/retrieval, the recall-rate (R) and precision-rate (P) may raise rapidly to fulfill requirements for business application. The synonyms dictionary within this essay mainly refers to the Chinese language, while in the case of English, "the WordNet" can be a reference to achieve the similar purpose. In the future, collaborative data exchange in business perspective will be studied, while the utility of the synonyms dictionary will be enhanced with Terms-similarity calculation, and the utility of semantic hierarchy structure will be enhanced with related clustering calculation.

Acknowledgements

This work was supported by Talent Fund Projectof Hefei University (No. 14RC08), the ministry of science and technology of P.R China Innovation Fund for technology based firms projects (No. 11C26213401181), Key Constructive Discipline of Hefei University (No. 2016xk05).

REFERENCES:


