

Article

Evaluate the Interoperability of Document Format: Based on Translation Practice of OOXML and UOF

Yaohu Lin ^{1,2,*}, Xuelian Lin ², Ning Li ³ and Yongmin Mu ³

¹ Department of Economics and Management, Beihang University, Beijing 100191, China

² School of Computer Science and Engineering, Beihang University, Beijing 100191, China;

E-Mail: linxl@buaa.edu.cn

³ Computer School, Beijing Information Science and Technology University, Beijing 100101, China;

E-Mails: ningli.ok@163.com (N.L.); yongminmu@163.com (Y.M.)

* Author to whom correspondence should be addressed; E-Mail: linyaohu@buaa.edu.cn;

Tel.: +86-134-363-55589; Fax: +86-10-8231-7832.

Academic Editor: Willy Susilo

Received: 23 January 2015 / Accepted: 19 March 2015 / Published: 27 March 2015

Abstract: Taking both OOXML and UOF standards as examples, we empirically evaluate the interoperability of office document formats from the view of translation practice. With the aim of covering the complete feature set of OOXML and UOF, a novel UOF-Open XML Translator is developed in this study. Thorough experiments demonstrate that our translator implements bidirectional conversion of 80.4% features perfectly and 9.9% features with acceptable discrepancy. Regarding the remaining 9.7% features, more efforts would be taken in future work.

Keywords: OOXML; OpenXML; UOF; interoperability; evaluate

1. Introduction

As the carrier of information and knowledge, document has been deep into every corner of social life. From personal letters, e-books to commercial contracts, government documents, representation and storage of document affects all our lives. In the 1990s, private binary document format was very common and the document was dependent on the software. At that time, the doc format of Microsoft became the defacto standard [1]. This causes a lot of compatibility and security issues for document information exchange, especially in the network environment which has different OS platforms.

Nowadays, governments, standards bodies and other organizations have found open standards for document formats can have more choice, lower cost and stimulate innovation [2]. This has emerged as a central issue for them. Open standard document standards, such as OpenDocument Format (ODF, ISO/IEC 26300:2006), Office Open XML (OOXML, ISO/IEC 29500:2008) [3] and Uniform Office Format (UOF, Chinese Government Standard GB/T20916-2007) are believed to provide a wealth of economic and technological benefits. Open document format has been accepted by more and more organizations and individuals, software such as OpenOffice, StarOffice, Google Docs support ODF standard well, Microsoft Office, Pages, ThinkFree Office support OOXML standard well and YOZO Office, King Office support UOF well.

The evaluation of the interoperability between different office document formats is important, especially in the new era of Big Data, because the interoperable way of documents processing is crucial for efficiency and compatibility. There are many studies in the office standard and there are many relative works for translating between different office standards, like OOXML-UOF Translator [4], UOF/ODF for word processing [5], Compare the Word Processing Part of OOXML and ODF [6], Evaluating the Interoperability of ODF and OOXML [2], *etc.* Some of these studies only focus on the standard, such as [1] and [7], which elaborate the history and their competition of OOXML and ODF, and these articles also show how the office standard affects the economy all over the world. Some of the interoperability research does not focus on the whole standard but parts of it, such as [6], which compares the OOXML and UOF standard based on the word processing part. The interoperability evaluating in [2] focuses on the theory, model study and the software support.

In this article, we empirically evaluate the interoperability of office document formats based on many years' document format translation projects and the document interoperability evaluating model research [8]. We take OOXML and UOF as examples to get the evaluating value of different office document formats according to comparing and analyzing the features of word processing, presentation and spreadsheet through the translation practice. From the results, we can see that all the office format standards can support the core features which people used very often well and the interoperability can be carried out easily, but there are discrepancies in some detail features especially in the enumeration type. Interoperability is difficult because different standards have their own definition in some features.

2. Background

2.1. Open Standard Document Format

2.1.1. OOXML

OOXML, also called OpenXML or Office Open XML, is an XML-based electronic file standard compress with ZIP. It includes the definition of file structure description, supporting word processing, spreadsheet and presentation.

OOXML became one of the standards of ECMA in Dec 2006 which number is ECMA-376. It passed the vote of international standard organization in April 2008, then announced the ISO/IEC 29500 international standard two months later.

Office Open XML file format is becoming Microsoft Office's default format after Microsoft Office 2007. Microsoft Office 2010 supports the reading operation of ECMA-376, reading and writing

operation of ISO/IEC 29500 Transitional. In the Microsoft Office 2013, it supports the reading and writing operation of ISO/IEC 29500 Strict [9].

2.1.2. UOF

The UOF—Uniform Office Format is an emerging standard, which is being developed by the Chinese Office Software Work Group (COSWG), led by the China Electronics Standard Institute (CESI), the Ministry of Industry and Information Technology (MIIT), major suppliers of Chinese office software suites, and other academic institutions.

China National Standardization Management Committee tabled the bill of national standard plan project in 2003 and it claimed some research institutes and companies to draft the UOF, which has intellectual property. UOF became the national office document recommendation standard in September 2009 [10]. Now the standard experiences the version of UOF 1.0, UOF 1.1 and UOF 2.0.

Uniform Office Format is an open standard for office applications, developed in China. It includes word processing, presentation, and spreadsheet modules, and is made up of GUI, API, and format specifications. The description of the document format uses XML, and is contained in a compressed file container. The UOF common contents are made up of Metadata, Styles, Hyperlink, Object set, User data, Digital Signature, and also include the convention of Measuring Unit, Anchor Represent Way and Linear Notation. Word processing, presentation and spreadsheets define the features in their part.

2.2. Interoperability

The standard of OOXML and UOF are all implemented based on XML technology and they all define and offer the implementation of office applications. These two standards support the compatibility of different office software and the function of transformation while there are many differences between them, which cause a lot of trouble regarding interoperability.

Interoperability is the capacity of exchanging and sharing data in different platform or programming language. Document interoperability refers to translate among different document standards [11]. This article takes OOXML and UOF as examples to consider the bidirectional interoperability capacity of different standards.

2.3. UOF-Open XML Translator Project

To improve the interoperability between OOXML and UOF in both directions, we have founded the UOF-Open XML working group. The working group analyzes the differences and similarities of these two standards and then implements the interoperability. After 7 years of effort, we have released seven versions of UOF-Open XML translator, test cases and test reports. They are:

- Version 5.0: provides Word processing, Spreadsheet and Presentation translations, including the translation of Open XML (ISO 29500 strict/transitional) to UOF 2.0 and the translation of UOF 2.0 to Open XML (ISO 29500 transitional). Performance and functionality enhancements over OpenXML/UOF Translator Version 4.1 have also been made in this project.
- Version 4.1: provides Word processing, Spreadsheet and Presentation translations, including the translation between Open XML (ISO 29500 transitional) and UOF 2.0 in bidirectional. More

performance and functionality enhancements over OpenXML/UOF Translator Version 4.0 have been made in this project.

- Version 4.0: provides Word processing, Spreadsheet and Presentation translations, including the translation between Open XML (ISO 29500 transitional) and UOF 2.0 in bidirectional. Performance and functionality enhancements over OpenXML/UOF Translator Version 3.0 have also been made in this project.
- Version 3.0: provides Word processing, Spreadsheet and Presentation translations, including the translation between Open XML (ISO 29500 transitional) and UOF 1.0/1.1 in bidirectional. Performance and functionality enhancements over OpenXML/UOF Translator Version 2.1 have also been made in this project.
- Version 2.1: provides Word processing, Spreadsheet and Presentation translations, including the translation between Open XML (ECMA 376) and UOF 1.0 in bidirectional and word processing translation between Open XML (ECMA 376) and UOF 1.1. Performance and functionality enhancements over OpenXML/UOF Translator Version 2.0 have also been made in this project.
- Version 2.0: provides Word processing, Spreadsheet and Presentation translations, including the translation between Open XML (ECMA 376) and UOF 1.0.
- Version 1.0: provides Word processing translation, only including the translation between Open XML (ECMA 376) and UOF 1.0.

The UOF-Open XML translator (also referred to as UOF Translator or OpenXML/UOF Translator) is an open source plugin. All the materials are published on the open source website [12], including the setup program, source code, design specification, test case, test report, and so on. All of these resources are opened to individual, company and institution, and everyone can download them free.

After installed the translator to the computer, there are several ways to use it. One of the usage modes is that you can see a menu in the explorer after install the translator successfully. People can use the context menu to translate the OOXML to UOF or translate UOF to OOXML, and it also supports batching translation. In this case, running the translator does not depend on the office software, even if you do not install any. In addition, we also develop an addin for Microsoft Office which users is able use our main translation program to open or save the UOF format file.

3. Interoperability Assessment and Test Method

The interoperability assessment methodology is that comparing and verifying all the features included in the standards. If the same feature in different standards can be fully equivalent, we say that it has completely interoperated. If there are only some parts of the feature that correspond, we say that it has partially interoperated. If the feature cannot correspond in different standards, then it cannot be interoperated in this point.

This research takes OOXML and UOF standards, for example to state the interoperability assessment of office document format standard according to the translation practice of OOXML and UOF.

3.1. Features

We divide word processing, spreadsheet and presentation into three feature levels in the interoperability assessment. First of all, we classify the standard into several parts that is the First Level. Then they are subdivided into a more detailed level called the Second Level. Finally, features are further subdivided to feature unit which is the Third Level. The detailed feature division is shown as Table 1.

Table 1. Feature Division of Document Format standard.

Standard Part	First Level feature	Second Level Feature	Third Level Feature
Word Processing	21	170	266
Spreadsheet	18	141	354
Presentation	23	174	387

The specific division is that word processing includes styles, revise, comment, index, region, *etc.*, which includes 21 features in the first level, and this level is divided into 170 sub-features in the second level, then divided into 266 feature units in the third level; the spreadsheet part includes rules, worktable setting, column setting, row setting, cell setting, *etc.*, which includes 18 features in the first level, and this level is divided into 141 sub-features in the second level, then divided into 354 feature units in the third level; the presentation part includes metadata, bookmark, hyperlink, style, *etc.*, 23 first features in its first level, 174 sub-features in its second level and 387 feature units in its third level.

3.2. Test Case

The test of interoperability for OOXML and UOF document formats based on a reference implementation approach and it covers all the features. For OOXML, most of the test documents are developed in Microsoft Office for windows. For UOF, most of the test documents are developed in YOZO Office and King Office.

We are trying to test all the features included in the OOXML and UOF standards. The newest test involves 106 test cases for word processing, 174 test cases for presentation and 207 test cases for spreadsheet. The test cases for word processing part tested 266 features, the presentation part tested 387 features and the spreadsheet part tested 354 features.

3.3. Interoperability Implementation

UOF-Open XML Translator uses the typical factory pattern to design the program which provides a unified interface. The program consists of pretreatment, main transform and post treatment. The pretreatment part is used to deal with the common preprocessing, such as read/write ZIP package, picture preprocessing, *etc.* The main transform part uses C# program to call the XSLT (Extensible Stylesheet Language Transformations) to complete most of the transformation, and some of the difficult transformation features which used XSLT is hard to carry out are completed in the post treatment.

The translator selects the right translation method according to the specific file (word processing/presentation/spreadsheet) when the translator is running. Most of the interoperability between OOXML and UOF standards are translated with XSLT which is a language for transforming XML documents into other XML documents. XSLT uses XPath (XML Path Language) to search for the

information in the XML file in the main transform, and then translates source XML tree (one of the office standards) to result XML tree (the other office standard).

Both the OOXML and UOF files are saved in the ZIP container. First of all, the translator analyzes the structure of original office document and builds the fundamental frame of target document, because different contents have different organization forms. All the UOF documents contain files of meta.xml, content.xml, rules.xml, styles.xml and uof.xml while they may contain extend.xml, graphics.xml, hyperlinks.xml, objectdata.xml and media files when the documents have the specific features. The OOXML components of word processing, presentation and spreadsheet consist completely different. Secondly, according to the structure of target document, translator gets all the needed information of every XML files from the original document and updates the content of target XML files. Finally, a translator compresses the generated XML files into ZIP form and updates the extension.

3.4. Limitation

The research in this article uses three feature level to compare OOXML and UOF standard, and then gets the interoperability degree value while does not consider which features are commonly used for users, and does not take the weight of every feature into account. Otherwise, the office software does not implement the whole standard and the office software itself will take into discrepancy during the interoperability, such as the YOZO Office does not support the text rotation in comment, 3D effects, etc. Some of the features of OOXML are implemented with VML in Microsoft Office.

4. Interoperability Assessment Practice and Results

4.1. Evaluating Model

For the given two office document format standards s' and s'' ,

$$E_s = Sim(F', F'') \quad (1)$$

where $s', s'' \in S$, S is the office document format set, F' is the feature set of s' and F'' is the feature set of s'' . E_s is the assessment result in Equation (1), $E_s \in [0,1]$. Sim is the evaluation function. For the specific feature, the evaluation function is

$$e = sim(f', f'') \quad (2)$$

where $e \in \{0,1\}$, f' is the element of F' and f'' is the element of F'' . When $e = 0$, it shows that the specific feature cannot be interoperability while $e = 1$ shows that it can be interoperability. According to Equations (1) and (2), the evaluation function also can be

$$E_s = \sum_1^n \frac{1}{n} * e_i, \quad i \in [1, n] \quad (3)$$

where n is the feature amount. Such as we divided the word processing part of OOXML and UOF into 266 features, then $n = 266$.

4.2. Interoperability Overview

In our newest UOF-Open XML Translator, we use 487 test cases to test the interoperability we have implemented. The result was shown as Table 2. Figures 1–3 are some of the translation effects, including word processing, presentation and spreadsheet.

Table 2. Translation result of Office Open XML (OOXML) and Uniform Office Format (UOF).

Standard Part	Feature with Discrepancy	Own Features	Total Feature	Full Equivalence	Discrepancy	Own
Word Processing	33	19	266	80.5%	12.4%	7.1%
Spreadsheet	26	55	387	79.1%	6.7%	14.2%
Presentation	37	28	354	81.6%	10.5%	7.9%

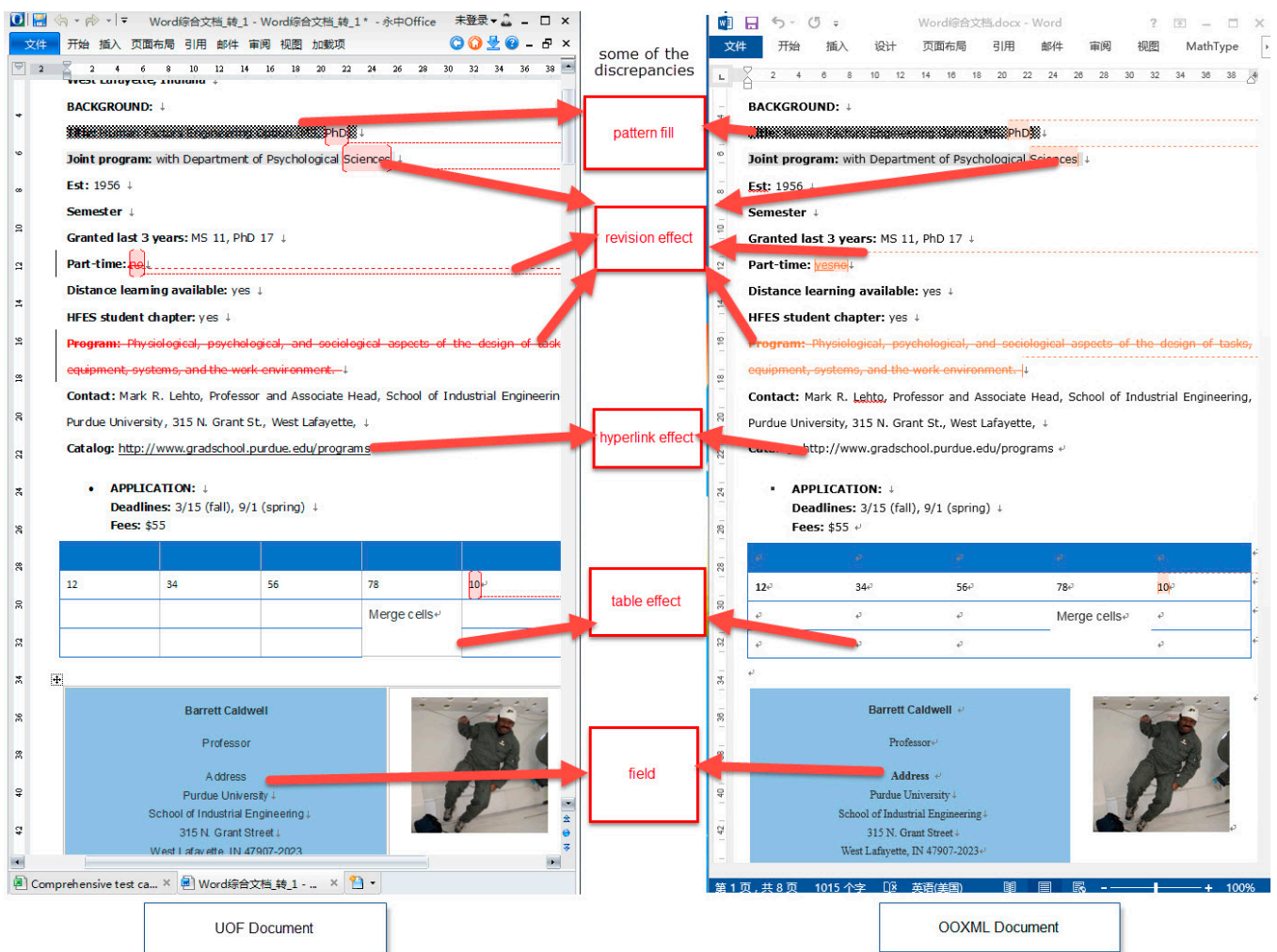


Figure 1. Word Processing Translation.

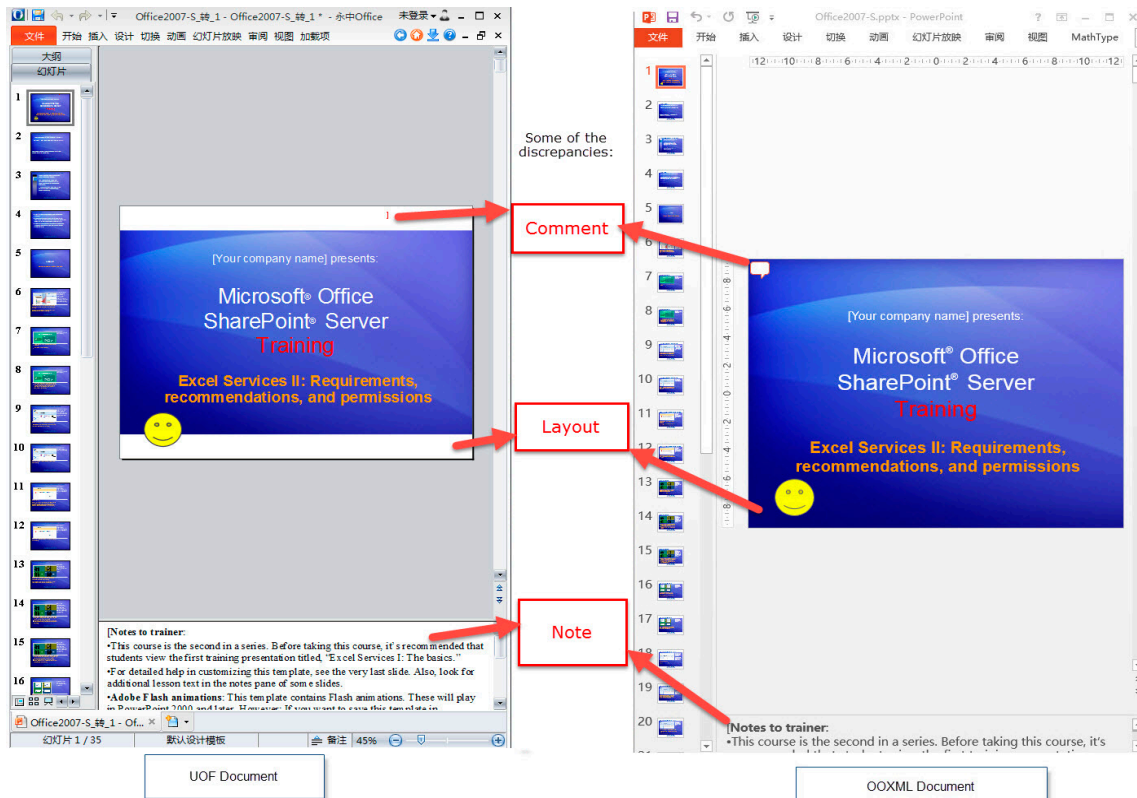


Figure 2. Presentation Translation.

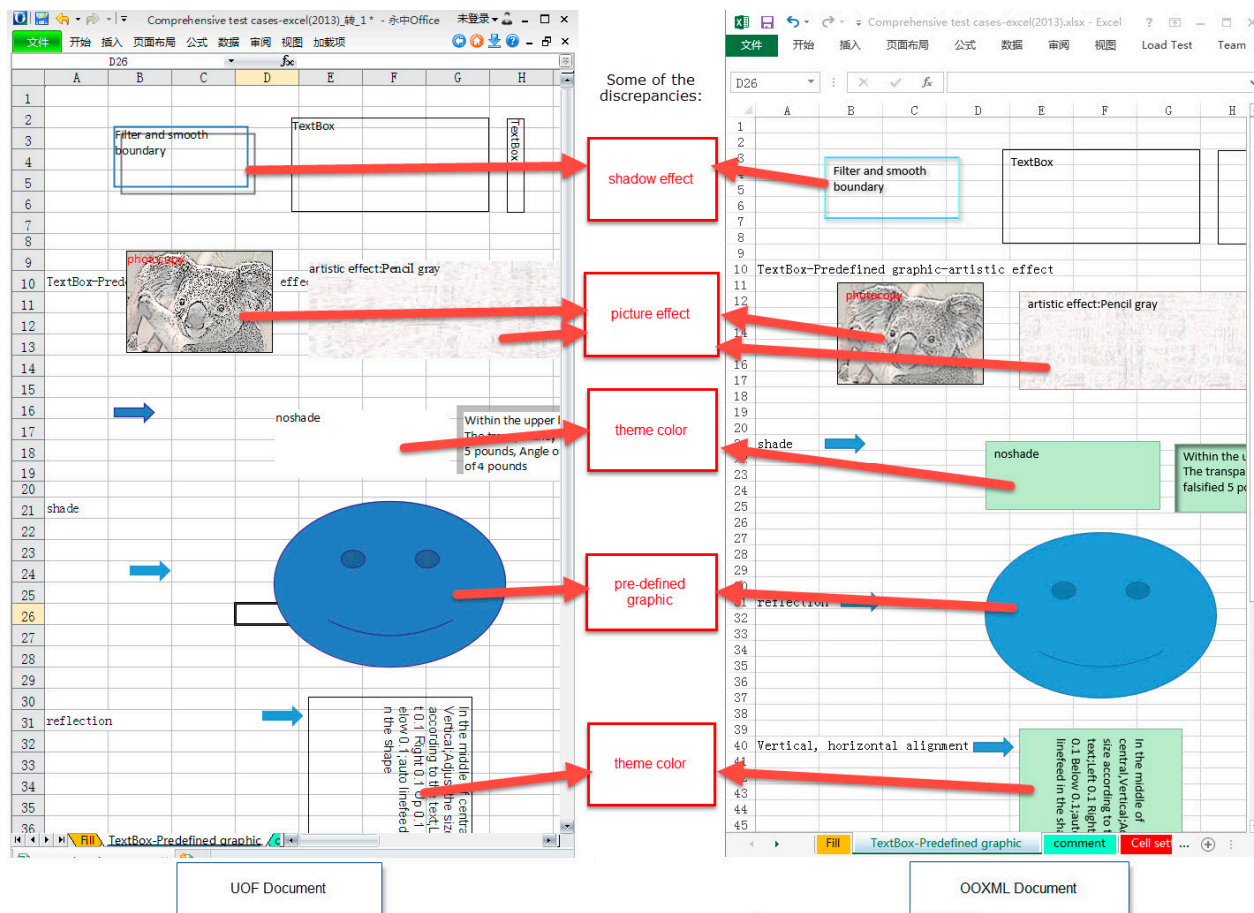


Figure 3. Spreadsheet Translation.

According to the evaluating model and the test result, we can see that the feature amount n of word processing is 266 and $E_s = \frac{266-33-19}{266} = 0.805$, which means about 80.5% of the features can be translated well in both direction. Moreover, there were 33 features in 266 which means that about 12.4% of the features can be translated with discrepancy while 7.1% of the features are not able to respond. Likewise, in the presentation part there are 79.1% of the features can be translated well in both directions, 6.7% of the features can be translated with discrepancy and 14.2% of the features that cannot be translated. In the spreadsheet part, 81.6% of the features can be translated well in both directions, 10.5% of the features can be translated with discrepancy and 7.9% of the features cannot correspond.

As a whole, the correspondence in the interoperability between OOXML and UOF, word processing part is the best which reaches 92.9%, the spreadsheet part is the second which reaches 92.1% and the presentation part is the worst which reaches 85.8%. However, in the features of full equivalence between OOXML and UOF, the spreadsheet part works best, then the word processing part. The presentation is the worst.

4.3. Core Features

From our translation practice, we find that the core features which people use commonly are supported well by different office document standard, such as: font, shape, size, color, bold, paragraph align, picture fill, background color, *etc.*, in the word processing; font, size, color, slides changing, common animation, *etc.*, in the presentation and font, color, size, common chart, *etc.*, in the spreadsheet. Moreover, the interoperability can be implemented easily.

4.4. Discrepancy Reason

Discrepancy exists while different office document formats are not exactly the same. We conclude that there are four reasons to result in the discrepancy from the translation practice.

The first case is the enumeration type. Different standards are hard to be exactly the same in the enumeration types, such as the rectangle in the pre-defined shape, there are nine shapes in the OOXML and only one shape in the UOF. Moreover, patterns fill, border type, animation switching, paper type, highlighted text, view and so on are belonging to this case.

The second case is that one of the standard does not define the specific feature while we can find the resemblance feature to match. In this case, no data lost will happen and the display effects will not be far from the original. The typical example is the 3D line chart, 3D area chart, *etc.* The OOXML defines these 3D charts, but these features do not exist in the UOF. In this case, we translate the 3D line chart to general line chart at the cost of 3D effect lost, but we save the data. In addition, the stock chart, chart in word processing, smartArt, section, layout and so on also belong to this situation.

The third case is that the feature relies on the software display. Some features like comment, superscript, shade, text overflow, *etc.*, have a large influence by the software. Frequently, the standards have the definition about these features in this case, but there are some differences in the visual effect.

The last case is that the feature is only defined in one of the standards. There are no similar features that can be found in the translation. Features such as region, measuring unit, access time, number of characters, slash header, formulary, hyperlink style and so on also belong to this case.

According to the newest UOF-Open XML Translator, the statistical result is shown as Table 3.

Table 3. Discrepancy distribution of OOXML and UOF.

Standard Part	Enumeration Discrepancy	Resemblance Discrepancy	Display Discrepancy	Own Discrepancy
Word Processing	13	5	4	30
Spreadsheet	12	6	5	39
Presentation	9	3	6	63

From the statistical result, we find the last case for which the feature only being defined in one of the standards is the main reason causing the discrepancy and it is also the main reason for interoperability (shown in Figure 4).

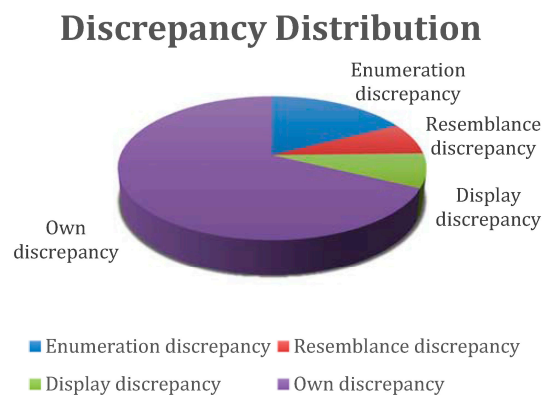


Figure 4. Discrepancy Distribution.

5. Conclusions

This article aims to evaluate the interoperability between different office document formats according to the translation practice. Some features can be translated in theory while there are some discrepancies in the translation implementation. Based on the OOXML and UOF translation program of several years, the results clearly indicate that more efforts should be taken to approach interoperability implementation.

The newest version of UOF-Open XML Translator shows that about 80.4% features can be translated between OOXML and UOF, about 9.9% of features can be translated with discrepancy and there are still about 9.7% features that we should study more closely.

In this study, we tested all the features of OOXML and UOF, but the test cases which are developed by Office software caused some problems when we verified the features. Because the files saved by Office software, such as Microsoft Office or YOZO, Office cannot ensure conformation to the standard. With more and more conformance and compatibility tests in office software by various organizations, the interoperability evaluation between different office document formats can be more precise.

Acknowledgments

Thanks to the support of Microsoft Corporation and Microsoft UOF working group, this work can go on. This research also relies on all graduate students of The Institute of Advanced Computing Technology of Beihang University and Beijing Information Science and Technology University UOF working group. All the test work are finished by Lenovo and Beijing Information Science and Technology University UOF Test Group.

Author Contributions

Conception and design: Yaohu Lin, Xuelian Lin; provision of study materials: Xuelian Lin, Ning Li and Yongmin Mu; collection and assembly of data: Yaohu Lin, Xuelina Lin, Ning Li and Yongmin Mu; data analysis and interpretation: Yaohu Lin, Xuelian Lin, Ning Li and Yongmin Mu; Manuscript writing: Yaohu Lin; All authors have read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Kosek, J. From the Office Document Format Battlefield. *Field Rep.* **2008**, *10*, 51–55.
2. Shah, R.; Kesan, J. Evaluating the Interoperability of Document Formats: ODF and OOXML as Examples. In Proceedings of the 2nd International Conference on Theory and Practice of Electronic Governance, Cairo, Egypt, 1–4 December 2008; pp. 219–225.
3. Li, N.; Wu, X.; Fang, C. The enlightenment to China of International Office document format of ODF and OOXML. *Tech. Hot Points* **2011**, *5*, 31–35.
4. Open XML/ODF Translator Add-ins for Office. Available online: <http://odf-converter.sourceforge.net> (accessed on 24 March 2015).
5. ODF-UOF Converter. Available online: <http://sourceforge.net/projectsodf-to-uof> (accessed on 24 March 2015).
6. Hou, X.; Li, N.; Yang, H. Comparison of Wordprocessing Document Format in OOXML and ODF. In Proceedings of the 2010 Sixth International Conference on Semantics, Knowledge and Grids, Beijing, China, 1–3 November 2010; pp. 297–300.
7. Blind, K. An economic analysis of standards competition: The example of the ISO ODF and OOXML standards. *Telecommun. Policy* **2011**, *35*, 373–381.
8. Li, N.; Liang, Q.; Hou, X.; Tian, Y. Document Interoperability Metric. *J. Beijing Inf. Sci. Technol. Univ.* **2011**, *26*, 6–13.
9. Office Open XML. Available online: <http://zh.wikipedia.org/wiki/OOXML> (accessed on 24 March 2015). (In Chinese)
10. UOF2.0. Available online: <http://baike.baidu.com/view/5116875.htm?fr=aladdin> (accessed on 24 March 2015). (In Chinese)
11. Interoperability. Available online: <http://baike.baidu.com/view/555117.htm?fr=aladdin> (accessed on 24 March 2015). (In Chinese)
12. UOF-Open XML translator. Available online: <http://uof-translator.sourceforge.net> (accessed on 24 March 2015).