

Doi: [10.15863/TAS](https://doi.org/10.15863/TAS)
**International Scientific Journal
Theoretical & Applied Science**
p-ISSN: 2308-4944 (print) e-ISSN: 2409-0085 (online)
Year: 2015 Issue: 02 Volume: 22
Published: 28.02.2015 <http://www.T-Science.org>

Alexandr Nikolayevich Shevtsov
candidate of technical sciences,
member of PILA (USA),
Department of «Mathematics»,
Deputy Director on Science of faculty of information
technologies, automation and telecommunications,
Taraz state University named after M.Kh. Dulati,
Kazakhstan
Shev_AlexXXXX@mail.ru

SECTION 2. Applied mathematics. Mathematical modeling.

THE COMPARISON ALGORITHM DOCUMENTS

Abstract: While handling the documents of a significant amount of is often a problem comparing them to match before and after editing. For tasks, this work describes a simple and effective code for Delphi Architect allows such an analysis of the documents in Word format and calculate the percentage of matches.

Key words: Delphi, Word, match, algorithm.

Language: Russian

Citation: Shevtsov AN (2015) THE COMPARISON ALGORITHM DOCUMENTS. ISJ Theoretical & Applied Science 02 (22): 16-20. doi: <http://dx.doi.org/10.15863/TAS.2015.02.22.4>

ОБ ОДНОМ АЛГОРИТМЕ СРАВНЕНИЯ ДОКУМЕНТОВ

Аннотация: При обработке документов значительного объема зачастую возникает проблема их сравнения на совпадения до и после редактирования. Для поставленной задачи, данная работа описывает простой и эффективный программный код на Delphi Architect позволяющий производить подобный анализ документов в формате Word и рассчитывать процент совпадений.

Ключевые слова: Delphi, Word, совпадение, алгоритм.

Сравнение документов достаточно актуальная проблема описываемая в различных источниках [1-8]. Сравнить документы вручную - занятие неблагодарное из-за слишком больших затрат времени и возможности ошибок, ведь не заметишь какую-то важную деталь при просмотре проще простого. Гораздо разумнее задачу сравнения файлов перепоручить компьютеру [1].

1. Сравнение непосредственно в Word.

В целом, в плане сравнения Word-документов все обстоит достаточно неудобно, хотя и без использования вспомогательных инструментов, в Word данная возможность надежно скрыта от чужих глаз, и, вероятно, не так много пользователей о ее существовании вообще догадываются [1]. Результат сравнения окажется представленным в новом документе, где слева будет отображен отрецензированный документ с учетом изменений, а справа (друг над другом) - исходный и измененный документы.

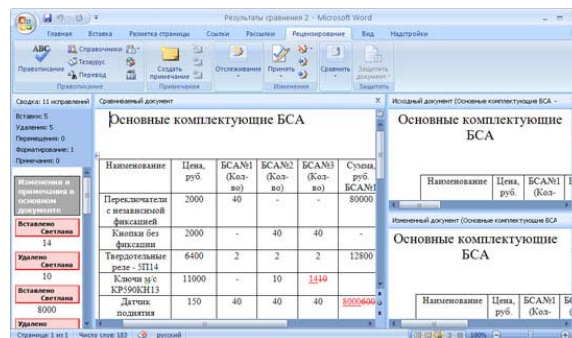


Рисунок 1 – Сравнение документов непосредственно в Word.

Теоретически, в Excel тоже возможно сравнение документов встроенными средствами, правда, только при работе в режиме фиксации изменений. Однако это неудобно, поскольку каждую из измененных ячеек придется просматривать, наводя на нее мышью, так как изменения, внесенные в документ, отображаются во всплывающих окошках (примерно таких, как обычные примечания) [1].

2. Сравнение в Compare Suite 7.0 компании AKS-Labs.

Достаточно удобный инструмент для быстрого сравнения текстовых файлов, документов MS Office, RTF-документов, файлов PDF, web-страниц (HTML), презентаций PowerPoint, бинарных и некоторых других типов файлов, а также файлов в ZIP- и RAR-архивах и на FTP-серверах. В случае сравнения листингов программ на ряде языков программирования (Object Pascal, HTML, C/C++, JavaScript, PHP и др.) предусмотрена подсветка синтаксиса. Имеется функционал для синхронизации текстовых файлов и сравнения папок вместе с подпапками [2].

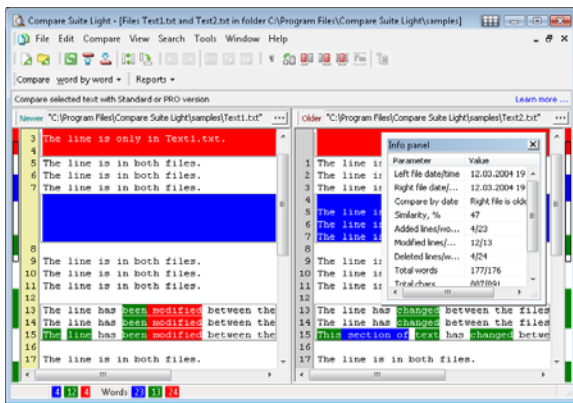


Рисунок 2 – Сравнение в Compare Suite.

Кроме этого комплексы сравнения также представлены следующими разработками:

3. Compare It! 4.1.2174 [3];
4. Beyond Compare 3.1.9 [4];
5. WinMerge 2.12.4 [5];
6. Active File Compare 2.0 [6];
7. Compare Spreadsheets for Excel 1.1.5 [7];
8. Excel Compare 2.3 [8];
9. xc Excel Compare Tools 1.0 [9];
10. Spreadsheet Compare 1.15 [10];

Приведенные программы обладая широким функционалом, неудобны в использовании, когда необходимо с минимумом совершаемых действий получить результат. А при смещении всего на 1 символ выдают абсолютно неверные результаты, так как основаны на побуквенном сравнении.

Как результат – они абсолютно неприменимы, например при анализе двух экзаменационных тестов на повторяющиеся вопросы. Необходима разработка более совершенных алгоритмов сравнения документов.

В основе нашего алгоритма будем использовать разбиение одного из документов на блоки и поблочное сравнение.

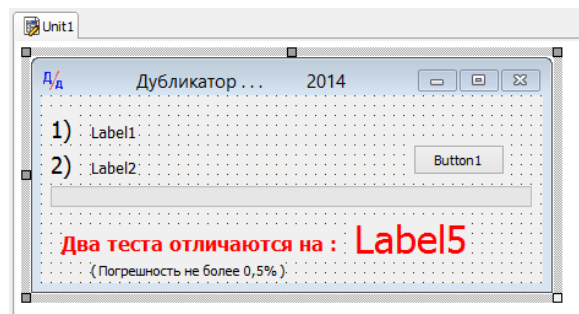


Рисунок 3 – Форма программы.

Разработаем алгоритмы и программный код на Delphi Architect. Загрузка документов будет осуществляться методом Drag&Drop, при последующей автоматической перезагрузке, и возможности не закрывая программу сразу же после анализа двух тестов перейти к анализу двух других. Действия пользователя сводятся к перетаскиванию двух файлов на форму программы, что значительно ускоряет и оптимизирует работу.

```

code: Delphi
unit Unit1;
interface

uses
    Winapi.Windows, Winapi.Messages,
    System.SysUtils, System.Variants, System.Classes,
    Vcl.Graphics,
    Vcl.Controls, Vcl.Forms, Vcl.Dialogs, Vcl.StdCtrls,

    ShellAPI, ComObj,
    Office_Tlb, Vcl.ComCtrls;

type
    TForm1 = class(TForm)
        Label1: TLabel;
    end;
    
```

```

Label2: TLabel;
Label3: TLabel;
Label4: TLabel;
Button1: TButton;
Memo1: TMemo;
Memo2: TMemo;
ProgressBar1: TProgressBar;
Label5: TLabel;
Label6: TLabel;
Label7: TLabel;
Label8: TLabel;
procedure FormCreate(Sender: TObject);
procedure Button1Click(Sender: TObject);
private
    { Private declarations }
protected
    
```

```
procedure WMDropFiles (var Msg: TMessage);
  message wm_DropFiles;
public
  { Public declarations }
end;

var
  Form1: TForm1;

s1,s2, F1,F2: string;
sd, n,z,i,zn,ii:integer;
x1,x2,p2,p1:double;
  Copies,Range: OLEVariant;
Background,collate:boolean;
PageType,Pages:string;
WordApp: OLEVariant;

implementation

{$R *.dfm}
function GetToken(aString, SepChar: string;
TokenNum: Byte): string; var Token: string;StrLen:
Byte;TNum: Byte;TEnd: Byte;
begin StrLen := Length(aString);TNum := 1;TEnd :=
StrLen; while ((TNum <= TokenNum) and (TEnd <>
0)) do begin TEnd := Pos(SepChar, aString); if
TEnd <> 0 then begin Token := Copy(aString, 1,
TEnd - 1); Delete(aString, 1, TEnd);
Inc(TNum); end else begin Token := aString;
end; end; if TNum >= TokenNum then begin
GetToken := Token; end else begin GetToken := "";
end;end;

function NumToken(aString, SepChar: string): Byte;
var RChar: Char; StrLen,TNum,TEnd: Byte;
begin if SepChar = '#' then begin RChar := '*' end
else begin RChar := '#' end; StrLen :=
Length(aString); TNum := 0;TEnd := StrLen;while
TEnd <> 0 do begin Inc(TNum); TEnd :=
Pos(SepChar, aString); if TEnd <> 0 then begin
aString[TEnd] := RChar; end; end;Result :=
TNum;end;

procedure TForm1.Button1Click(Sender: TObject);
begin
label5.Caption:="";
if not(fileexists(F1)) then exit;
if not(fileexists(F2)) then exit;

WordApp:=CreateOleObject('Word.Application');
WordApp.Visible:=false;
//открываем документ
WordApp.Documents.Open(f1);

memo1.Text:=WordApp.ActiveDocument.Range.Te
xt;
```

```
WordApp.ActiveDocument.Close(true,EmptyParam,
EmptyParam);
WordApp.Documents.Open(f2);

memo2.Text:=WordApp.ActiveDocument.Range.Te
xt;

WordApp.ActiveDocument.Close(true,EmptyParam,
EmptyParam);

sd:=50; // коэффициент сравнения

s1:=memo1.Text;
s2:=memo2.Text;

zn:=trunc(length(s1)/sd);
progressbar1.Position:=0;
progressbar1.Max:=zn;

z:=0;

for I := 1 to zn do
begin
if pos(copy(s1,i*sd,sd),s2)>0 then inc(z);
progressbar1.Position:=i;
application.ProcessMessages;
end;

x1:=int((zn-z)/zn*10000)/100;
ii:=zn;

s1:=memo2.Text;
s2:=memo1.Text;

zn:=trunc(length(s1)/sd);
progressbar1.Position:=0;
progressbar1.Max:=ii+zn;

label7.Caption:=floattostr(ii)+' '+floattostr(zn);

p1:=300/ii;
p2:=300/zn;

if p1>p2 then p2:=p1;
label8.Caption:="( Погрешность не более
'+floattostr(int(p2*100)/100)+' % )";

z:=0;

for I := 1 to zn do
begin
if pos(copy(s1,i*sd,sd),s2)>0 then inc(z);
progressbar1.Position:=ii+i;
application.ProcessMessages;
end;
x2:=int((zn-z)/zn*10000)/100;
```

```
if x1<x2 then x2:=x1;

label5.Caption:=floattostr(x2)+' %';
// а потом и word
WordApp.Quit(false,EmptyParam,EmptyParam) ;
end;

procedure TForm1.FormCreate(Sender: TObject);
begin
label7.Caption:="";
label8.Caption:="";

DragAcceptFiles(Handle,True);
n:=1;
label1.Caption:='___';
label2.Caption:='___';
label5.Caption:="";
end;

procedure TForm1.WMDropFiles(var Msg:
TMessage);
Var
Filename: array[0..256] of char;
begin

DragQueryFile(THandle(Msg.WParam),0,Filename,
SizeOf(Filename));
// Теперь в переменной Filename будет
находиться путь // к перетаскиваемому файлу.
Далее вы можете выполнять с этим файлом, зная
// его путь, все что угодно.
//Например: Загрузить его в Мемо//
Memo1.lines.loadfromfile(Filename);
//Сообщаем об окончании претаскивания

case n of
1:begin label2.caption:='___';
label1.caption:=extractfilename(Filename);
F1:=Filename;
inc(n);end;
2:begin
label2.caption:=extractfilename(Filename);F2:=File
Name;n:=1;Button1.Click;end;

end;

DragFinish(THandle(Msg.WParam));
end;

end.
```

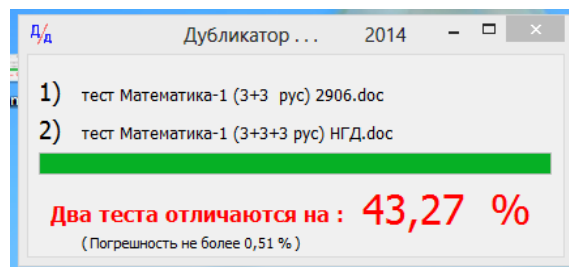
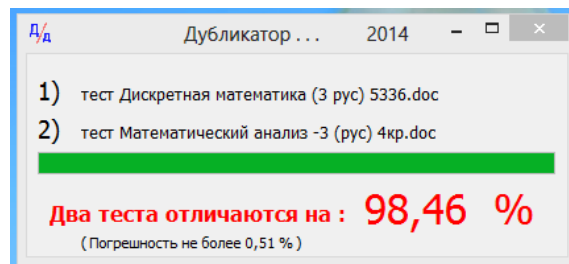


Рисунок 4 – Результаты анализа тестов.

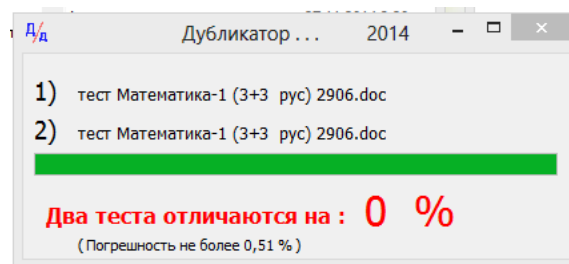


Рисунок 5 – Сравнение одинаковых тестов.

Разработанные алгоритмы и программа внедрена в производство с 2014 г. и прошла апробацию в центре тестирования ТарГУ им.М.Х.Дулати.

References:

1. Shlahina S (2010) Tool for the rapid comparison of documents.. Available:

<http://www.3dnews.ru/585650>

(Accessed:

27.02.2015).

Impact Factor ISRA (India) = 1.344
Impact Factor ISI (Dubai, UAE) = 0.829
based on International Citation Report (ICR)

Impact Factor JIF = 1.500
Impact Factor GIF (Australia) = 0.356
Impact Factor SIS (USA) = 0.438

2. (2014) Compare Suite Light – Freeware Compare Tool. Available: <http://www.comparesuite.com/free-file-comparison.htm> (Accessed: 27.02.2015).
3. (2015) Compare It! Available: <http://www.grigsoft.com/> (Accessed: 27.02.2015).
4. (2015) Beyond Compare 3.1.9 Available: <http://scootersoftware.com> (Accessed: 27.02.2015).
5. (2015) WinMerge 2.12.4 Available: <http://winmerge.org/> (Accessed: 27.02.2015).
6. (2015) Active File Compare 2.0 Available: <http://www.formulasoft.ru/active-file-compare.html> (Accessed: 27.02.2015).
7. (2015) Compare Spreadsheets for Excel 1.1.5 Available: https://www.mapilab.com/ru/excel/compare_spreadsheets/ (Accessed: 27.02.2015).
8. (2015) Excel Compare 2.3 Available: <http://www.formulasoft.com> (Accessed: 27.02.2015).
9. (2015) xc Excel Compare Tools 1.0 Available: <http://code.google.com/p/xcexcelcompare/downloads/detail?name=xc.Excel.CompareTools.v1.0.en.rar&can=2&q=>
10. (2015) Spreadsheet Compare 1.15 Available: <http://www.lumadownload.com/ru/software/326282/spreadsheet+compare+1.26> (Accessed: 27.02.2015).

