## Text mining solution to support the evaluation of research grant proposals

Detection of plagiarism, scientific overlap, double funding, resubmission.
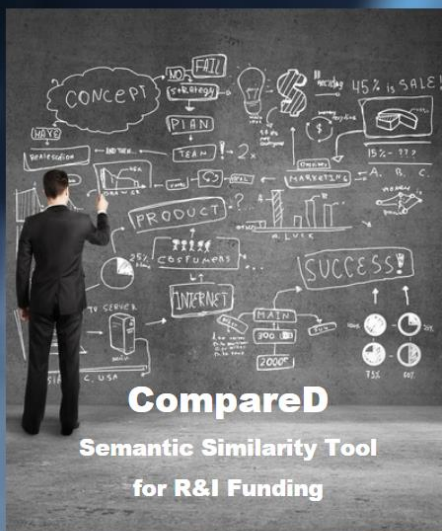
Funded by DG Digit under ISA² programme

1 years pilot project

Scaling up

       -> budget available for 2019

       -> more testing by users => new features

       -> more data

              -> which data?

                     -> funding agencies in member states (reach out)

                     -> open access data (e.g. NIH)

       -> link to data hub initiative?

# Compared



**Login to CompareD**

Submit Your Credentials Here:

Username:

Password:

☑ I accept the terms of use

Submit

**CompareD**

**Semantic Similarity Tool**

**for R&I Funding**

# Compared

## CompareD
**Semantic Similarity Tool for R&I Funding**

Logout

## What Do You Want To Do Next ?

Start New Matching Session    See History    Manage Users

# CompareD

**Semantic Similarity Tool for R&I Funding**

## Step 1. Upload Documents Or Copy/Paste Text for Comparing With Various Reference Documents

A. Please select the file(s) to perform the comparison.

Add File(s)

Selected Files

B. Or Copy/Paste below the desired text to compare

Add Text

Copy/Pasted Texts

Back    Next

# Compared

## CompareD
**Semantic Similarity Tool for R&I Funding**

Logout

### Step 1. Upload Documents Or Copy/Paste Text for Comparing With Various Reference Documents

A. Please select the file(s) to perform the comparison.

Add File(s)

**Selected Files**

| | |
|---|---|
| BioElectricSurface Final Project Summary.pdf | 430.47 Kb ✔ ⊗ |
| test fp7.docx | 14.08 Kb ✔ ⊗ |

B. Or Copy/Paste below the desired text to compare

Add Text

**Copy/Pasted Texts**

SUNSET (http://www.sunset-project.eu) is a three-year research & development project part of the European Commission's Seventh Framework programme Smart Cities & Sustainability under DG Connect (http://ec.europa.eu/dgs/connect/en/content/smart-cities). SUNSET started in February 2011 and has ended February 2014. SUNSET develops and evaluates a set of innovative services that use Smartphone technology, social networks and incentives to encourage people to travel more sustainably in urban environments. The project's objective is to increase personal mobility and at the same time reduce congestion, increase safety, and protect the environment. The SUNSET project uses a human-centred approach to achieve its objectives stimulating people to change their individual travelling behaviour. To influence behaviour, we developed and exploited a Smartphone application named tripzoom featuring challenges and rewards to move smarter. Moreover, we tailor and personalise these incentives by means of automatically measuring actual travel behaviour of the Smartphone user.. This is a personalised, multimodal coaching approach to traffic and mobility management, based on rewarding good behaviour. SUNSET is an initiative of a consortium of nine public and private partners from four different European countries with a total budget of 4.1 million euro. SUNSET combines technical with social research creating new services for sustainable travel and evaluation of these services in real life settings of the cities Enschede (NL), Gothenburg (SE) and Leeds (UK).

Remove

Back    Next

# Compared

## CompareD
### Semantic Similarity Tool for R&I Funding

## Step 2. Review Input Files And Texts

Please note that only texts in English can be at this moment matched.

| | File Or Copy Pasted Text | Language |
|---|---|---|
| 1 | File (size=440806): BioElectricSurface Final Project Summary.pdf | English |
| 2 | File (size=14414): test fp7.docx | English |
| 3 | Text: SUNSET (http://www.sunset-project.eu) is a three-year research & development project part of the European Commission's Seventh Framework programme Smart Cities & Sustainability under DG Connect (http://ec.europa.eu/dgs/connect/en/content/smart-cities). SUNSET started in Fe | English |

Back    Next

European
Commission

# CompareD
### Semantic Similarity Tool for R&I Funding

Logout

## Step 3. Select the Comparison Parameters. These Parameters Will Impact the Number and Quality of Obtained Results.

### Select the Reference Sources to Compare With

| | |
|---|---|
| Cordis ☑ | Cordis is the Community Research and Development Information Service. It is the European Commission's primary public repository and portal to disseminate information on all EU-funded research projects and their results in the broadest sense. |
| Patstat ☐ | |
| Scopus ☐ | PATSTAT contains bibliographical data relating to more than 100 million patent documents from leading industrialised and developing countries. |
| Wikipedia ☐ | Scopus is Elsevier's abstract and citation database launched in 2004. Scopus covers nearly 36,377 titles (22,794 active titles and 13,583 Inactive titles) from approximately 11,678 publishers, of which 34,346 are peer-reviewed journals in top-level subject fields: life sciences, social sciences, physical sciences and health sciences. It covers three types of sources: book series, journals, and trade journals. |

### Select the Types of Words Matching

○ Only Nouns
○ Nouns and Adjectives
● Nouns, Adjectives & Adverbs
○ Nouns, Verbs, Adjectives & Adverbs

Use this section to select the Parts-of-Speech based on which text comparison is made. For example, if you select the "Only Nouns" option, the match shall be made via cosine similarity between texts which where stripped out of all other parts of speech except nouns.

Probably "Only Nouns" is the best option for comparing scientific documents, but the other options are also useful.

### Select the Similarity % Threshold for Detection

40 %

Similarity threshold influences the number of results obtained. The lower this threshold is the more results will be obtained.

Similarity is the Cosine Similarity computed by Term Frequency - Inverse Document Frequency.

### Select If to Perform Plagiarism Verifications Or Not

Perform Plagiarism Verifications ☐

Similarity threshold influences the number of results obtained. The lower this threshold is the more results will be obtained.

Similarity is the Cosine Similarity computed by Term Frequency - Inverse Document Frequency.

### Date Interval for Reference Documents

1997 - 2019

Please note that plagiarism verifications are very time consuming and could prolong significantly the time required to perform the comparisons.

Back     Next

# Compared

European Commission

## CompareD
### Semantic Similarity Tool for R&I Funding

Logou

**1/3: SUNSET (http://www.sunset-project.eu) is a**

There were no matchings found with similarity above the selected

Details

### Number of Documents Per Similarity Range

5
4
3
2
1

15-20%  20-30%  30-40%  40-50%  50-60%  60-70%  70-80%  80-90%  90-100%
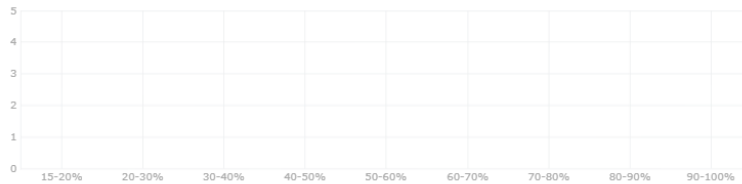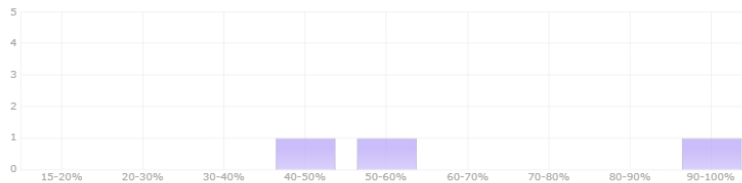
### Number of references With Common Words

- 0-10 Common Words: 0 References
- 10-30 Common Words: 0 References
- 30-50 Common Words: 0 References
- 50-70 Common Words: 0 References
- 70-100 Common Words: 0 References
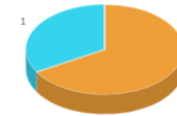- > 100 Common Words: 0 References

**2/3: test fp7.docx**

Number of results (Similarity > Selected threshold): 3

Details

### Number of Documents Per Similarity Range

5
4
3
2
1
0

15-20%  20-30%  30-40%  40-50%  50-60%  60-70%  70-80%  80-90%  90-100%
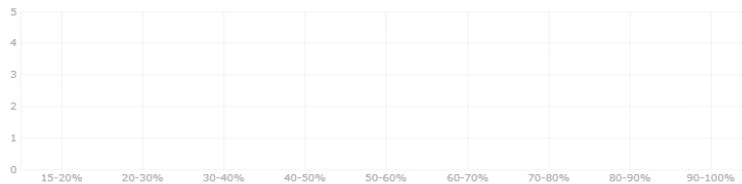
### Number of references With Common Words

- 0-10 Common Words: 0 References
- 10-30 Common Words: 2 References
- 30-50 Common Words: 0 References
- 50-70 Common Words: 0 References
- 70-100 Common Words: 1 References
- > 100 Common Words: 0 References

**3/3: BioElectricSurface Final Project Summary.pdf**

There were no matchings found with similarity above the selected

Details

### Number of Documents Per Similarity Range

5
4
3
2
1
0

15-20%  20-30%  30-40%  40-50%  50-60%  60-70%  70-80%  80-90%  90-100%

### Number of references With Common Words

- 0-10 Common Words: 0 References
- 10-30 Common Words: 0 References
- 30-50 Common Words: 0 References
- 50-70 Common Words: 0 References
- 70-100 Common Words: 0 References
- > 100 Common Words: 0 References

Back To Main Page    Start New Matching Session    Comparison Parameters

# Compared

## CompareD
### Semantic Similarity Tool for R&I Funding

### TYPE_SELECTED_FILE

| Reference GUID | Reference Title | Similarity | Year | Common Words | Words Only In Reference | Words Only In Text | Details | Reference |
|---|---|---|---|---|---|---|---|---|
| fp7_604251 | Reinforced Bioresorbable Biomaterials for Therapeutic Drug Eluting Stents | 100.00000000000 | 2014 | 84 | 0 | 0 | Show | View |
| fp7_251681 | Development of a Nitric Oxide Releasing Stent for Treatment of Coronary Artery Disease | 51.661177293328 | 2010 | 14 | 66 | 70 | Show | View |
| fp7_278313 | Biodegradable Magnetic Stent for Coronary Artery Luminal Regeneration | 42.614351834940 | 2012 | 10 | 74 | 74 | Show | View |

Back     Main Page     Start New Matching Session

# Compared

Logout

## Text To Be Matched - Common Words Are in Red

## Reference GUID: fp7_251681

Reference Title: Development of a Nitric Oxide Releasing Stent for Treatment of Coronary Artery Disease

Active therapeutic biodegradable and **biocompatible** materials are **highly** in demand. These are required for the production of medicinal products in a variety of areas including implant technology, tissue engineering, **drug** delivery and wound **healing**. Within implant technology such **biomaterials** can be used for dental, bone and cardiovascular implants. Tailored mechanical **properties**, **biocompatibility** and degradation rate is the key to the development for a specific implant. **stents** are tubular type implants that are deployed most commonly to recover the shape of narrowed arterial segments. Although, the clinical use of **stents** is widespread, they cause adverse responses including **inflammation**, in-**stent restenosis** and **thrombosis**. Endothelialisation of the **stent** greatly reduces these adverse reactions. In contrast to permanent **stent** there is great attraction in the notion of a biodegradable **stent** that recovers and maintains arterial shape and then gradually disappears and avoids further complications. In this multi-institution & disciplinary SME focussed project we will aim to provide the technological framework that leads to the production of reinforced polymeric **biomaterials** tailored towards **stent** manufacturing without adverse **effects**. Both natural and synthetic polymers will be produced and used. These will be reinforced and functionalised using a variety of techniques. Controlled delivery of suitable positive additives including antimitotic factors will be aimed for and their **release** monitored. These **highly** functionalised active **biomaterials** will be characterised thoroughly for material **properties**, **biocompatibility**, rate of biodegradation and used for the production of ideal **stents**. These will be characterized thoroughly leading to preclinical validation. All required production and manufacturing guidelines will be followed.

Coronary artery disease is one of the leading causes of death worldwide. Significant advances in treating coronary artery disease have been made over the last decade with the introduction of percutaneous coronary intervention (PCI). PCI involves the placement of **stents** within narrowed arteries relieving the symptoms of angina and in some cases increasing life expectancy. The two major problems with currently available **stent** designs are **restenosis** and late **thrombosis**. The former is more common in uncoated **stents** and was addressed with the introduction of **drug** releasing **stents**. However, concerns have been raised about the increased incidence of **stent thrombosis** with these **drug** releasing **stents**. An innovative design of **stent** utilising titanium oxynitride (Titan 2) was the first of the bioactive **stents** which compared favourably in efficacy to the **drug** coated **stents** but with less re-narrowing than with the uncoated **stents**. The Titan 2 uses the fundamental **biocompatibility** of titanium to reduce **inflammation** and promote **healing** after stenting. Nitrous oxide (NO) is an important regulator of endothelial cell function. We will design a bioactive titanium oxynitride coated **stent** which will **release** NO after **stent** implantation. It is proposed that this will have beneficial **effects** to endothelial function after stenting and possibly reduce the incidence of re-narrowing and **stent thrombosis**. The new metal **stent** will be tested in vitro using primary human monocytes and platelets and using a human derived endothelial cell line. We aim to achieve excellent **biomaterial properties** as applied to coronary artery stenting, a **highly biocompatible** surface and to demonstrate **release** of NO after **stent** deployment. To achieve this we will combine the knowledge of our industrial, medical and scientific partners to develop a new coronary **stent** which will be of significant benefit to patients with coronary artery disease both in the European Union and worldwide.

## A. Common Words (Lemmas):

biocompatibility,biocompatible,biomaterial,drug,effect,healing,highly,inflammation,property,release,restenosis,stent,stents,thrombosis

## B. Words Only In Input Text:

active,additive,adverse,antimitotic,area,arterial,attraction,biodegradable,biodegradation,bone,cardiovascular,clinical,commonly,complication,contrast,degradation,delivery,demand,dental,development,disciplinary,endothelialisation,engineering,factor,framework,functionalised,gradually,great,greatly,guideline,ideal,implant,institution,key,manufacturing,material,mechanical,medicinal,multi,natural,notion,permanent,polymer,polymeric,positive,preclinical,product,production,project,rate,reaction,recovers,response,segment,shape,sme,specific,suitable,synthetic,technique,technological,technology,therapeutic,tissue,tubular,type,validation,variety,widespread,wound

## C. Words Only In Reference Document:

advance,angina,artery,beneficial,benefit,bioactive,case,cell,coated,common,concern,coronary,death,decade,deployment,design,disease,efficacy,endothelial,excellent,expectancy,favourably,function,fundamental,human,implantation,important,incidence,industrial,innovative,intervention,introduction,knowledge,late,life,line,major,medical,metal,monocyte,narrowing,nitrous,oxide,oxynitride,partner,patient,pci,percutaneous,placement,platelet,possibly,primary,problem,regulator,scientific,significant,stenting,surface,symptom,titan,titanium,uncoated,union,utilising,vitro,worldwide

Back

## Current status

Final meeting pilot phase 4$^{th}$ December: **decision to scale up or not** (funding ok).

**Scaling-up project**

> -> more data

> -> users in funding agencies

> -> link to data hub project

> -> improve tool with new user requirements

**Question to Science Europe**

Is this needed?

How to scale it up? Promotion?

Data?