Embedding a Creativity Support Tool within Computer Graphics Research

Yalemisew Abgaz[1], Diarmuid P. O'Donoghue[1], Donny Hurley[1], Horacio Saggion[2], Francesco Ronzano[2]

Abstract. We describe the Dr Inventor creativity support tool that aims to support and even enhance the creativity of active research scientists, by discovering un-noticed analogical similarities between publications. The tool combines text processing, lexical analysis and computational cognitive modeling to find comparisons with the greatest potential for a creative impact on the system users. A multi-year corpus of publications is used to drive the creativity of the system, with a central graph matching algorithm being adapted to identify the best analogy between any pair of papers. Dr Inventor has been developed for use by computer graphics researchers, with a particular focus on publications from the SIGGRAPH conference series and it uses this context in three main ways. Firstly, the pragmatic context of creativity support requires the identification of comparisons that are unlike pre-existing information. Secondly, the suggested inferences are assessed for quality within the context of a corpus of graphics publications. Finally, expert users from this discipline were asked to identify the qualities of greatest concern to them, which then guided the subsequent evaluation task.

1. INTRODUCTION

Creativity is a highly valued human ability, lying at the heart of many advances in scientific thinking and processes. Reasoning with the use of analogical comparisons [1] is a well-known explanation for many instances of scientific creativity and can also be a driver of scientific creativity [2]. Creativity support tools (CST) [3] aim to facilitate users in their efforts to produce some creative output. Dr Inventor [4] is a CST focused on creativity within scientific reasoning, helping in the creation of novel information that is useful to some scientific community.

We view the creative process as being composed of distinct sub-tasks, with Dr Inventor to perform some tasks while the user retains overall responsibility for the creative outcomes. Dr Inventor assumes responsibility for identifying high quality analogical comparisons between scientific publications (related to its application domain, computer graphics), based on a computational model [5] of the human ability of reasoning using analogies. Dr Inventor adopts a Big Data perspective towards creative inspiration, by exploiting the wide availability of academic documents for use as sources of inspiration for Dr Inventor's users. The user is then responsibility for ultimately evaluating and either using the presented analogy – or rejecting it as a false or fruitless comparison.

For example, many papers in computer graphics addressing the problem of cloth simulation use "thin plate equations" to simulate the look and behavior of clothes. But using these equations is based on an analogy between a piece of cloth and a thin metallic plate. The problem of modelling clothes is the *target/problem* while the metallic place is called the *source*. Even if such comparisons may seem obvious once they are presented, generating novel and useful analogies is a very difficult and challenging problem.

In this paper we present a novel combination of lexical and semantic processing with a computational analogy model, aimed at discovering novel and useful analogies between publications. Section two provides an overview of creativity and how it is supported by the process of thinking analogically. Section three describes the text processing pipeline and the subsequent generation of a semantic graph structure. Section four describes the core analogy model and its computational metrics. Section five then describes the document corpus and user studies that evaluated the effectiveness of the identified analogies.

2. ANALOGICAL COMPARISONS IN CREATIVE SCIENTIFIC REASONING

Creativity is a highly valued human ability and can be seen as a form of self-generated thought that produces new and useful knowledge, which makes subsequent reasoning more effective. We focus on creativity driven by bisociations [6] between disparate concepts, relying on the well-studied cognitive process of reasoning through the use of analogical comparisons.

Analogies pervade our understanding, particularly of complex or abstract concepts such as time [7]. Analogies involve comparisons between dissimilar objects, but the degree of semantic difference between the *source* and *target* analogs can vary greatly. A *target* from one area of computer graphics may be compared to a different area of computer graphics (often called "near analogies") or to politics or cooking ("far analogies"). Semantically far analogs have long been associated with more innovative and challenging comparisons. Notably, scientific revolutions [8] are strongly associated with these semantically distant comparisons.

While Dr Inventor is not yet aiming at identifying creative analogies that might revolutionize some scientific discipline, it does hope to uncover latent analogies that might drive scientists'

[1] Department of Computer Science, Maynooth University, Co..Kildare, Ireland
[2] Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain.

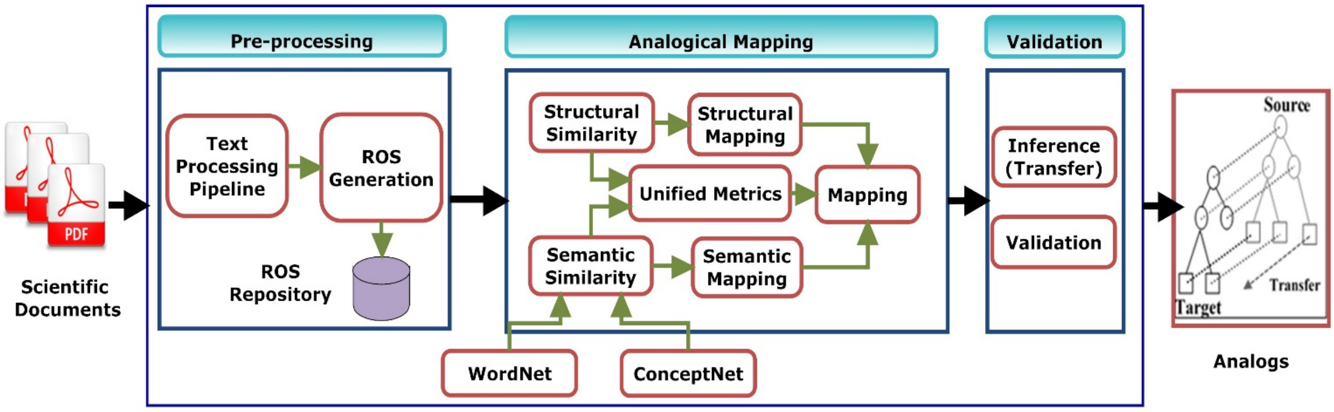

Figure 1. Dr. Inventor's Analogically Blended Creativity Framework

creativity. The role of analogies in scientific reasoning can be easily overlooked. A study of 16 one-hour meetings held across four different biological laboratories, identified the use of over 99 distinct analogies [9]. The majority of these analogies involved comparisons between semantically near items, such as comparisons between similar organisms or parts of organisms. This paper explores potentially creative "near" analogies between graphics publications.

[9] found that "far" analogies were often used to formulate a new hypothesis, using comparisons between an organism and (say) physics or even politics. Far analogies have also been shown to promote relational thinking [10], highlighting deep analogous similarity and overcoming any superficial similarities that may exist.

2.1. Computational Creativity

Computational creativity is a new discipline that aims to emulate human creativity, producing outputs that possess the central traits of creativity: *novelty* and *quality* (or *usefulness*) [11]. [12] demonstrated that a computational model of analogy is capable of generating many creative scientific analogies, but this work was limited by its reliance on hand-coded data. The approach adopted in this paper overcomes that limitation by sourcing all data directly from published documents, utilizing only machine-based processing of the original problem data. Dr Inventor forms and evaluates all of its analogical comparisons from the "raw" publications [13] using its novel combination of lexical and semantic processing.

2.2. Boosting Creativity with the Dr Inventor CST

We present the Dr Inventor CST (Figure 1) that aims to foster the creativity of practising scientists based on a cognitive computation model to simulate the generation of many analogies. From the results generated by our model, we choose the best analogical comparisons that offer a (potentially creative) interpretation of a given problem paper to ignite the scientist's creativity. Dr Inventor takes a descriptive computational model of the analogical reasoning process and uses it to predict those analogies that will have the greatest impact on its users' creativity.

The following factors are intended to help identify those analogies with the greatest creative potential:

- an extensive corpus with many candidate sources with which to re-interpret any given target problem
- metrics focused on identifying "good" analogies with creative potential
- persistence in exploring many analogies

Our CST addresses several of the challenges that are known to inhibit peoples' creativity:

- problem fixation and being entrenched in one view of a problem [14]
- memory limitations [15] and access to potentially useful information
- [16] showed that people do not notice analogies even when they are presented to them, but Dr Inventor can exhaustively explore all analogies [17] [18]

Additionally, our computational model enables us to quantify some metrics to help identify creative analogies by

- quantifying the level of pre-existing similarity between papers (using metrics based on the WordNet lexical database) and
- estimating the relative importance of pre-existing similarity and inferences for creative analogizing.

This paper explores the related challenges of developing and assessing the outputs of a CST within the specialized context of computer graphics research. We avail of experts in computer graphics to assist in this evaluation process. The major components of the tool are discussed in detail in Section 3 and 4.

2.3. Creativity in Computer Graphics Context

To ascertain the importance of creativity in the context of researchers in computer graphics, two surveys were undertaken. The first survey sought the opinions of practising researchers within this discipline as to the level of importance they placed on creativity when reviewing conference or journal papers. Respondents were asked for their opinion on the value they placed on creativity when reviewing papers. Three statements were rated by respondents:

1. Creativity is important when reviewing paper.
2. I can assess the level of creativity in a paper.
3. I can compare the levels of creativity between two papers

We believe the results shown in Figure 2 provide strong support for the importance of creativity in scientific research. Over 75% of respondents either "Strongly agreed" or "Agreed" that creativity is important when reviewing a paper. Additionally we infer that creativity is important to the research underlying such publications. Around 80% of respondents said they are able to assess the level of creativity of a paper (presumably in part by detecting differences was previously read papers). Only the last question attracted a small level of disagreement, suggesting that comparing the level of creativity between two papers may sometimes be quite challenging.

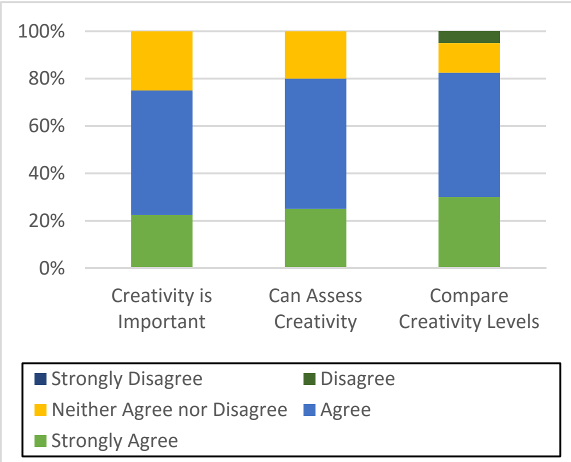

Figure 2: Do authors and reviewers of publications believe that creativity is important in a paper

Buoyed by this support for creativity within scientific research, we focused on specific metrics for use in evaluating the outputs of Dr Inventor. The SPECS standard [19] identified 14 independent components of general creativity, this encompassed creativity from diverse disciplines like the culinary arts, poetry, painting and architecture, with components like *emotion and self-expression* and *spontaneous and subconscious processing*. Thus, a survey was undertaken to identify the SPECS components of greatest relevance to scientific creativity and computer graphics researchers, with 34 researchers rating each quality on a 5-point Likert scale. The three qualities identified as most relevant to scientific creativity (by researchers in computer science) were as follows:

1 This is a novel or unexpected comparison (M=4.3, sd=0.73)

2 This comparison is potentially useful and recognizes gaps in current research (M=4.1, sd=0.83)

3 This comparison challenges the norms in this discipline. (M=3.8, sd=0.99)

Later, we shall see how these three qualities were used by respondents to evaluate the analogies developed by Dr Inventor.

3. SYNTACTIC AND SEMANTIC PROCESSING

3.1. Dr Inventor Text Mining Framework

The semantic analysis of the research articles and the extraction of *subject-verb-object* triples from the text of papers is supported by the Dr Inventor Framework [20] (DRI Framework), a pipeline of text-mining modules. The DRI Framework is distributed as a stand-alone Java library[2] that exposes an API to trigger the analysis of articles as well as to easily retrieve the results. In particular, the Framework defines a data model [21] of scientific publication properly structured to accommodate and conveniently expose the result of the analyses performed over a paper.

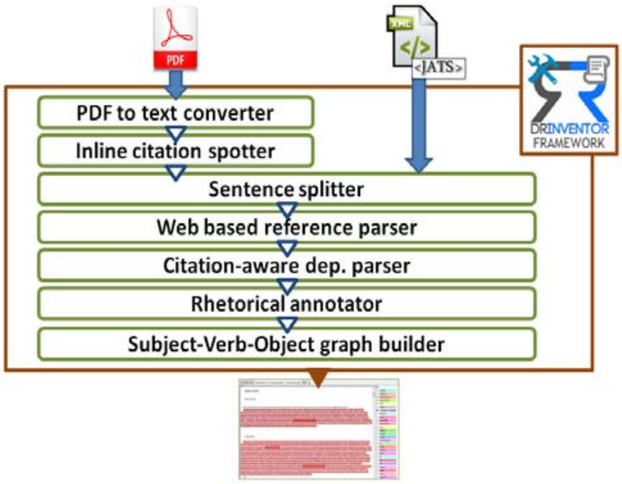

Figure 3: Architecture of the Dr Inventor Text Processing Framework

Figure 3 provides an overview of the core scientific text mining modules of the DRI Framework. Since most scientific publications are available in PDF format, the *PDF to text converter* processes PDF articles by invoking the PDFX online Web service[3] [22]: papers are converted into XML documents by identifying core structural elements including the title, the abstract, the hierarchy of sections and the bibliographic entries. This step can be by-passed if the article is available as *JATS***.** Citations are identified by the *Inline citation spotter* relying on a set of high coverage regular expressions and heuristics. Sentence boundaries in the documents are identified by a *Sentence Splitter* specifically customized to the idiosyncrasies of scientific discourse. The bibliographic entries identified in the article are enriched by means of the *Web based reference parser* by accessing external Web services including Bibsonomy[4], CrossRef[5] and FreeCite[6]. In order to obtain syntactic dependencies between words in each sentence, a *Citation-aware dependency parser* builds the dependency tree of the sentences using [23] which we have customized so as to correctly deal with in-line citations. Since the rhetorical role of a sentence in a scientific document is important for information extraction and other scientific content analysis activities, a trainable logistic regression *Rhetorical classifier* was developed which assigns to each sentence of a paper a rhetorical category (i.e. Background, Approach, Challenge, Outcome and Future Work). The classifier is trained on the Dr Inventor Multi-layered Corpus[7] of Computer Graphics papers, manually annotated in the context of the Dr Inventor Project [18]. This corpus was used to train the classifier.

By relying on the output of the dependency parser, the *Subject-Verb-Object graph builder* extracts from the contents of a paper Subject-Verb-Object triples as shown in Figure 4. These triples constitute the core structure of the ROS graph that is mined in order to spot similar papers and analogies among the contents of publications.

[2] The Dr. Inventor Text Mining Framework Java library can up the the IP be downloaded at: http://backingdata.org/dri/library/
[3] http://pdfx.cs.man.ac.uk/
[4] http://www.bibsonomy.org/help/doc/api.html
[5] http://search.crossref.org/help/api
[6] http://freecite.library.brown.edu/
[7] http://sempub.taln.upf.edu/dricorpus/

Even if not explicitly shown in Figure 3, the Dr Inventor Framework also supports the generation of extractive summaries of publications by implementing several approaches to select the most relevant sentences to be included in the summary [24] which can be used to select triples occurring in the most relevant parts of a document.

3.2. ROS-graph Generation

The analogy system does not work directly on the publications but instead uses a graph-centered representation based on the text extraction. These graphs are called Research Object Skeleton (ROS) graphs.

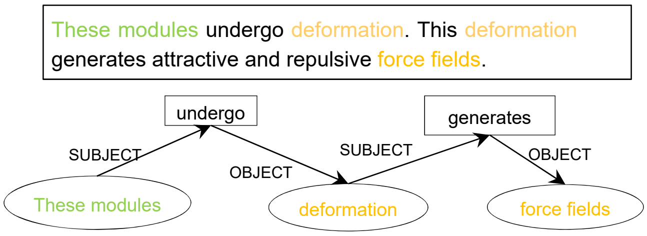

Figure 4: Subject-Verb-Object triples generated by the graph builder from two sentences

The ROS graphs have at the core of their structure the Noun-Verb-Noun type of relations (or Concept-Relation-Concept) enabling the application of Structure Mapping Theory [25] of analogy formation. While the core of the graph is the triple structure, the graph format chosen can have relationships between relations, i.e. second-order relations or causal relationships between nodes. These graphs are a form of *attributed relational graph* where nodes haves the attribute of "type" (i.e. noun, verb, causal). Among the additional attributes added to each node we consider the rhetorical category associated to the sentence in which the node occurs, extracted by means of the text processing pipeline and represented as an ontology-based semantic annotation [26]. This enables the creation of sub-graphs where analysis can be made on particular chosen categories of the publication. Dr Inventor relies on, for storage, the graph database Neo4j[8] which uses attributed relational graphs as its representation – making it highly suitable for our purposes.

The ROS is constructed by considering the dependency tree derived from each sentence in the publication. As in [27] a set of rules is applied to these trees, generating connected triples of nouns and verbs. One of the key properties of the ROS graphs is that multiple mentions of the same concept are uniquely represented. This is done either from the co-reference resolution of the text mining framework or by simply joining nodes that have the same word. Relation nodes, i.e. the verbs, can appear multiple times in the ROS. These constructed ROS graphs enable the steps of the analogy process and the mapping between different publications.

4. ANALOGY GENERATION AND METRICS

Analogy generation involves a mapping between the ROS of a selected target paper and the available source papers. The mapping pairs are then evaluated using a number of metrics and the best analogies are presented for evaluation by users.

4.1. ROS Mapping

Finding creative analogy requires exploration of many unsuccessful comparisons before discovering any useful analogy. Because of the high computational cost of performing retrieval, mapping and evaluation on a great many comparisons, computational efficiency was a primary concern – especially in the design of this central mapping phase.

Following Gentner's structure mapping theory [25], we generate the mapping between the source and target graphs. Our mapping involves structural mapping based on the graph structures and semantic mapping based on the semantics represented by the individual nodes and edges of the graphs. We also utilize mapping rules and constraints discussed in [28] distinctly incorporating both structural mapping and semantic aspects into the mapping process.

Generating the inter-ROS mapping is primarily driven by structure – that is, driven by any similarities between the topologies of the two ROS graphs. Thus, topology serves as a hard constraint on the space of possible mappings that is considered by Dr Inventor. However, when the structure of the two ROSs indicate multiple alternative solutions, we use semantic similarity to guide development of the preferred mapping. Thus, semantics are used as a soft constraint (or a preference constraint) on the mapping process, choosing between alternative mappings when different interpretations are available.

4.1.1. *Structural Mapping*

Our structural mapping is based on graph structure and conceptual structure. Graph structure focuses on identifying isomorphic graphs, while conceptual structure addresses the conceptual similarity between the nodes and edges that are to be paired by the mapping process [16, 29]. Specifically the objective of our structural mapping is to find the largest isomorphic subgraphs of a target paper in a source papers. For our specific purposes, let S be the set of all nodes in the source ROS graph $G_S=(S, E_S)$, let T be the set of all nodes in the target ROS graph $G_T= (T, E_T)$ and let $M = \{(S_i, T_i)| \, S_i \in S, T_i \in T, S_i \text{ is mapped to } T_i\}$ be the set of mappings between the source graph and the target graph. A mapping $M \subset S \times T$ is said to be an isomorphism iff M is a bijective function that preserves the branch structure of the graphs. And M is said to be the best analogical mapping if: 1) M is an isomorphism between a subgraph of G_T and subgraph of G_S, 2) M is the largest subgraph and, 3) M has the highest semantic similarity between its pairs.

We consider three constraints to guide structural mapping. The first constraint is defined on the types of nodes. A pair of nodes should have the same conceptual category to be a candidate of structural mapping. This means, "nouns" only map to "nouns" and "verbs" map only to "verbs". The second constraint is defined on the type of the edges. For two edges to be considered candidates, their corresponding nodes should satisfy the first constraint. We included the commutativity of relation (verb) nodes in a graph. If we consider a commutative relationship like (x adjacent y) and noting that this is equivalent to (y adjacent x), we allow such commutative relations to map more flexibly than non-commutative relations. The third

[8] http://www.neo4j.com

constraint focuses on the degree of the mapping nodes. The degree of a candidate node of the source graph should be at least greater than the degree of the target node. This allows us to find isomorphic subgraphs. In addition to these constraints, the traditional definition of structural mapping [25] holds true for this discussion.

Our structural mapping is implemented using a customized version of graph matching algorithm called VF2 [30]. The customization introduced the above constraints to preserve the properties of analogy mapping.

4.1.2. *Semantic Mapping*

Semantic mapping is an aspect of the mapping process that favours the generation of mappings that place a small cognitive workload on the Dr Inventor users – favouring semantically "simple" analogies whenever these are possible. This preference constraint is based on the similarity of the meaning of the words represented by each node in the ROS. Our semantic mapping utilizes the Lin similarity measure [31], which is based on WordNet [32], to calculate the similarity between source nodes and target nodes of similar type. These semantic similarity values are used during the computation and the selection phase of candidate pairs to be included in *M*. A pair with higher similarity score is selected and expanded first whenever we encounter two or more feasible candidate pairs. Thus, semantic mapping ensures a higher semantic similarity between the words represented by the mapping nodes of the isomorphic subgraph.

4.1.3. *Lexico-Semantic Features*

The text processing pipeline, ROS generation and analogy formation were largely developed as separate components, a number of features of each were aimed at maximizing the analogies that could be formed and their creative potential. The text processing pipeline and its dependency parser aimed to maximize the number of complete subject-verb-object triples, so that the rich and highly connected ROS graphs could be generated to form large rich mappings. The automated identification of the rhetorical category of sentences allows Dr Inventor to identify analogies between different parts of publications. This paper focuses on analogies formed between papers, each represented by its (lexical) "Abstract" and the rhetorical category of "Background".

We readily acknowledge that Dr Inventor does not have a deep understanding of the analogies it generates. Thus it could not be used to reliably create a new document from any of its discovered analogies for addition to its corpus. Therefore, it has not yet reached the level of being able to support the kind of *self-sustaining* computational creativity discussed in [33].

4.1.4. *Inference and Validation*

Inferences suggested by the analogy are modeled through the CWSG – Copy With Substitution and Generation [34] – a form of inference generation through of pattern completion. Dr Inventor ensures that all inferences are "grounded" in the mapping to ensure no spurious inferences are generated. While this paper explored analogies *only* between graphics publications and the resulting inference should (generally) be plausible combinations of source and target information, we report on some initial work aimed at validating inferences. Each inference is in the form of a triple (S V O), with each term arising in either the source or the target paper. A necessary step before evaluating Dr Inventor using publications outside the discipline of computer graphics, is to validate the inferences by detecting spurious combinations of S, V and O that may inadvertently arise.

Inference validation is one as the main mechanisms utilizing the graphics context and we explored several approaches to validating inferences. Firstly, inferences may be validated through comparison with existing triples in the Dr Inventor corpus by identifying a pre-existing instance in the Neo4j database. For less familiar triples an N-Gram model was developed to calculate the likelihood of combinations of S, V and O.

$$P(s,v,o) = P(s|\langle start\rangle)\, P(v|s)\, P(o|v)\, P(\langle end\rangle|o)$$

However, the N-Gram approach would be greatly hampered by zero probabilities arising from the novel (i.e. creative) combinations that Dr Inventor seeks. We explored additive smoothing [35], Good-Turing smoothing [36] and synonym substitution. Finding quality synonyms for the computer graphics context proved challenging an initial testing indicated that ConceptNet was not appropriate to validate graphics inferences. For this paper we focused on the WordsAPI provided by an online service[9].

4.2. Metrics

Once we generate the mappings between each source and target ROS, we further analyse the result to compute some metrics related to analogical similarity. This involves independent assessment of the semantic and structural factors involved in similarity. We then used a unified metric computed by multiplying structural similarity by semantic similarity. For measuring structural similarity we used Jaccard's coefficient [37]. The coefficient is used to measure the similarity between two finite sets, *A* and *B*. It is defined as:

$$J(A,B) = |A \cap B|/|A \cup B| \ = |A \cap B|/(|A| + |B| - |A \cap B|) \quad (1)$$

The Jaccard's coefficient gives a value of 1 if the A and B are structurally identical and yields 0 if there is no commonality between the two sets. Recall that $M = \{(S_i, T_i)|\ S_i \in S, T_i \in T, S_i \text{ is mapped to } T_i\}$. The Jaccard's coefficient for two graphs is then $J(S,T)$ where M is effectively $S \cap T$. Therefore, $J(S,T) = 0$, if there is no mapping between the two ROSs and $J(S,T) = 1$, if the two ROSs are structurally identical. Jaccard's coefficient gives a good estimation of *how much* of the two graphs have been mapped. For measuring semantic similarity between a pair of words, different approaches are suggested by research [38].

4.2.1. *WordNet based metrics*

The Lin metric returns value between 0 and 1 and has a readily accessible API. The overall semantic similarity of the mapping pairs is given by the average semantic similarity of the pairs in M, i.e.

$$\mathrm{SemS}(M) = \frac{\sum_{i=1}^{m} \mathrm{Lin}(S_i, T_i)}{m}, \quad (2)$$

where $m = |M|$ is the size of the mapping. Novel words not known within WordNet were not included in these calculations. A unified metric is computed as the product of the structural similarity and the semantic similarity. Unified Analogy similarity (AS) metrics is given as:

$$\mathrm{AS}(S,T) = J(S,T) \times \mathrm{SemS}(M) \quad (3)$$

---
[9] http://www.mashape.com

To support the identification of analogous papers, we use the Lin metric to calculate independent levels of *relational similarity* – between mapped verbs and *conceptual similarity* between mapped nouns. This allows Dr Inventor to identify mappings with high relational similarity but low conceptual similarity, although there is no agreed definition of low and high.

An additional metric quantifies the number of inferences that are mandated by each analogical comparison, as modeled through a simple pattern-completion process based on the inter-ROS mapping. More inferences may indicate a comparison highlights something new about the target problem and we expect (at least) some of these inferences to be useful and meaningful if we adapt them from the source to the target paper.

5. EVALUATION OF GENERATED ANALOGIES BY EXPERTS

We present the setup of the experiment and evaluation results. To evaluate the performance of the system, we run our tool using a computer graphics collection of papers. Experts from computer graphics domain evaluated the results of the system. We ask the users to rate the analogs based on selected properties of creative systems identified by SPECS [19] and collect both quantitative and qualitative feedback. We present the results below.

5.1. Experimental Conditions

5.1.1. Datasets – computer graphics corpus

A corpus of computer graphics publications formed the basis for this evaluation, consisting of publications from the ACM Special Interest Group on Computer Graphics and Interactive Techniques (SIGGRAPH) conference – the top-ranked conference on computer graphics according to Microsoft Academic Search. The corpus contained 957 papers from the proceedings of SIGGRAPH between 2002 and 2011. Papers ranged from 6 to 12 double column pages. Each paper of the corpus was processed by the DRI Framework, thus identifying sentences together with their rhetorical category (challenge, background, approach, outcome, etc.). A typical ROS graph contains an average of 997 nodes (median=1013, mode=1041 and SD=±265).

Ten *target* papers were selected using a simple random sampling technique, with their titles being listed in Table 1. For the experiments reported in this paper we considered only the triples generated from the abstract and from its sentences classified as background (rhetorical category) of each paper. This reduced the burden on evaluators by allowing them to focus on a subset of the paper (highlighted by a customized pdf viewer). Second, this reduced the size of the graphs, greatly expediting the computational process of finding the largest mapping.

Dr Inventor was then used to generate all possible analogies for each target, using all 957 papers in the corpus as potential sources. From the resulting 957 analogical comparisons, the best source paper was selected for each target using the metric described in section 4.2.

5.1.2. Overview of Respondents

The outputs of the system were evaluated by 14 active researchers working in different areas of computer graphics. Their professional level includes postgraduate students (9), postdoctoral researchers (2), senior lecturers (2) and professors (1). The gender distribution is female (4) and male (10). The evaluation task was preceded by users watching a training video and the entire evaluation task was completed over two days. Postgraduate evaluators were compensated for their participation in this evaluation task.

5.1.3. Evaluation procedure

Before the evaluation, the respondents were presented with a short introductory video outlining analogy and analogy based comparisons. Then they were introduced to the Dr Inventor system and their evaluation task.

Each analog pair of papers was presented and evaluated in turn. Users had access to the pdf version of the papers, including a highlighting of the sentences from the rhetorical "background" category. Users also were able to see the terms that had been placed in correspondence by the analogical mapping process, to help them better understand the presented analogy.

The system also allowed the users to browse the ROS graph thanks to an interactive visualization. The system further allowed users to navigate to/from the source and the target papers to the ROS visualization to find the original text where the mappings occurred. After spending sufficient time studying the analogs, users then gave their feedback on each analogous pair of papers.

5.2. Expert Ratings for the 10 good Analogies

The 14 researchers rated the 10 analogs, found by the Dr Inventor system, (No 1 to 10) for the 3 qualities discussed in Section 2.3 using a 5 point Likert scale [1-5]. While the number of respondents may appear small, each evaluation required reading two graphics publications and interaction with Dr Inventor system to explore the similarities using the visualization tools. 14 users evaluated 10 analogies each (reading 20 papers) with each analogy evaluation taking around 45 minutes. Thus our detailed evaluation represented around 110 person hours of work (or almost 14 8-hour work days).

Table 1. List of SIGGRAPH paper titles that formed the best analogies

No	Target Paper	Creative Source Paper
1	*Linear Combination of Transformations*	*Gaussian KD-Trees for Fast High-Dimensional Filtering*
2	*Curve Skeleton Extraction from Incomplete Point Cloud*	*Fast Bilateral Filtering for the Display of High-Dynamic-Range Images*
3	*Deforming Meshes that Split and Merge*	*Near-Regular Texture Analysis and Manipulation*
4	*Rotational Symmetry Field Design on Surfaces*	*Subdivision shading*
5	*3D Modeling with Silhouettes*	*Invertible Motion Blur in Video*
6	*Converting 3D Furniture Models to Fabricatable Parts and Connectors*	*Multi-Aperture Photography*
7	*Physical Reproduction of Materials with Specified Subsurface Scattering*	*Enrichment Textures for Detailed Cutting of Shells*

8	*Unstructured Video-Based Rendering: Interactive Exploration of Casually Captured Videos*	*Popup: Automatic Paper Architectures from 3D Models*
9	*Robust Treatment of Collisions, Contact and Friction for Cloth Animation*	*Inverse Shade Trees for Non-Parametric Material Representation and Editing*
10	*Real-Time Hand-Tracking with a Color Glove*	*Direct-to-Indirect Transfer for Cinematic Relighting*

Table 1 lists the titles of the source and target papers involved in each of the 10 analogies generated by Dr Inventor. Table 2 lists the computational metrics derived from each of these 10 analogical comparisons, grouped under the "Metrics" heading. Additionally, the average ratings awarded to each of these analogies under the three categories (novel useful and challenge) is also listed, grouped under the "Ratings" heading. The analogies in table 1 and also in table 2 have been ordered on descending values of the overall user ratings.

Table 2. Metrics and expert evaluations for the 10 generated analogies

	Metrics					Ratings				
Number	Rel Sim	Con Sim	M Ratio	Number of Inferences	Analogical Similarity	Novel	Useful	Challenge	Avg Rating	LSA
1	0.79	0.37	0.72	16	0.24	4.5	3.7	4.0	**4.07**	0.4
2	0.80	0.37	0.58	12	0.25	3.9	3.2	3.4	**3.48**	0.5
3	0.67	0.56	0.65	1	0.30	3.8	3.3	3.3	**3.44**	0.6
4	0.62	0.48	0.50	5	0.10	3.8	3.4	3.2	**3.44**	0.4
5	0.62	0.48	0.70	2	0.24	3.9	3.1	3.3	**3.43**	0.7
6	0.75	0.38	0.54	9	0.22	3.8	2.8	3.3	**3.30**	0.2
7	0.66	0.37	0.60	5	0.04	3.5	3.5	2.8	**3.28**	0.7
8	0.71	0.41	0.71	6	0.24	3.5	2.9	2.8	**3.08**	0.6
9	0.66	0.53	0.59	6	0.11	3.8	2.5	2.6	**2.97**	0.8
10	0.65	0.51	0.66	3	0.26	3.8	2.5	2.5	**2.92**	0.7

The top ranked analogy pair (No 1 in Table 1) has average user ratings of 4.46, 3.73 and 4.00 for the three qualities respectively and has an overall average of 4.06. The second ranked analogy pair (no 2) has a rating of 3.88, 3.05, and 3.33 with average rating of 3.42. However, the overall correlation between the analogical similarity and the user ratings is not strong. This leads to a further investigation of the proposed analogy metrics.

We do not expect all analogies generated by Dr Inventor to be rated highly for novelty, usefulness and challenging the norms. Figure 5 compares the ratings given to the best analogy with the average ratings awarded to all these analogies. The best analogy received higher than average ratings on each of the three qualities.

Looking particularly at the (computational) metrics for the top two analogies, an interesting pattern emerges. Firstly, these two analogies have the highest relational similarity (*RelSim* in Table 2) and the lowest conceptual similarity (*ConSim* in Table 2). These two qualities are the essential hallmarks of good analogical comparisons [1]. The larger *ConSim* scores indicate a difference in the nominals being discussed and are a strong indication that the anaology involves information arising from different research contexts – suggesting the source is document likely to be overlooked by a researcher. Additionally, these two analogies generated the largest number of inferences. A Pearson product-moment correlation of 0.608 was found between the number of inferences and the user ratings of each analogy, supporting importance of inferences to quality of analogies. Interestingly, the metrics for the two best comparisons displayed the classical hallmarks of good analogical comparisons is seen as strong support for both our approach and our computational model.

We also highlight that Dr Inventor's finds similarities that are

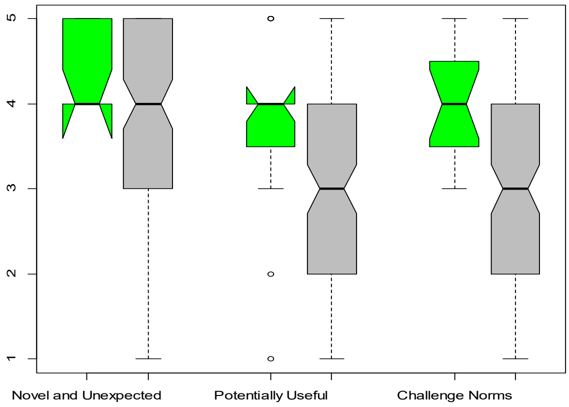

Figure 5. The ratings for the best and average analogies for each of the three qualities of creativity

different to other techniques by comparison to Latent Semantic Analysis (LSA), which has been used in previous work on analogy identification [39]. The LSA model was set to make its comparisons in document-to-document mode, using the first 300 factors of the "General Reading up to 1st year college" training set, which was used as a loose reflection of the linguistic exposure of the respondents (the majority of whom were postgraduate students).

The final column in Table 2 illustrates the (LSA) [40] score between analogous papers, using the lexical Abstract with rhetorical Background of each paper. The Pearson product-moment correlation between the *analogy score* and the LSA score was 0.1948 indicating that Dr Inventor is identifying documents that are quite dissimilar to those identified by LSA (noting that the corpus used for these results concerned *only* publications from SIGGRAPH). Similarly, the Pearson product-moment correlation between the user ratings and the LSA score was -0.523 indicating that Dr Inventor's users and LSA are identifying very different types of similarity between documents.

5.3. Qualitative Feedback

As well as quantitative feedback, two senior professors further identified their favorite analogs from the 10 generated pairs. The first user favored analogy number 1 (Table 2). This comparison suggested interesting relations. The subtopics of the two papers (interaction versus image, photography animation and collision), their year of publication (2002 and 2009 respectively) and the problems the two papers tried to solve were surprisingly different. The technique adopted by the target paper could be used in the context of the source paper, suggesting that "manipulations applied to *filters* can be applied to *matrices* and vice versa "*leading to a few possible research questions*".

The second user favored analogy number 2 (Table 1). The target paper covers topics such as modeling and point cloud whereas the source focuses on topics such as image processing and photography.

Here the target paper is published in 2009 whereas the source was published in 2002. The first paper addresses the problem of incomplete data during 3D laser scan, where the point cloud data representing the object contains large *holes* where the laser did not scan. The second paper addresses the problem of poor management of light for under/over exposed *areas* in a photographs. The respondent found that the suggested mappings are useful to recognize the technique used in one could be used in the other regardless of the different problem areas the two papers tackle. One evaluator was particularly interested in the mappings between "*hole*" and "*area*" and also between "*region*" and "*window*" (see Table 3). This professor noted that these two terms are generally used very differently and that thinking of one as being like the other was highly unusual and thought-provoking - despite the fact that the WordNet metrics did not show them to be particularly different. This analogy suggested that techniques described in the source paper could be used to effectively solve the problem of the target paper. Based on this analogy, the user suggested new ideas such as the use of the technique in the source paper to reconstruct hidden information for missing video data, facial expression, motion capture, recovery of 3D scan, X-ray etc.

Table 3. Excerpts from the mapping of analogy 2.

Source Word	Target Word	Sim Score	Source Word	Target Word	Sim Score
use	Utilize	1	outlier	source	0
function	information	0.350	**area**	**hole**	**0.419**
domain	Key	0.342	relate	to_compute	0.505
use	Be	0.774	weight	mesh	0.458
do_address	to_handle	1	**window**	**region**	**0.390**

One unexpected result of the evaluation is that some users found inspirations from the target to the source - while we only expected users to gain inspirations from the source to the target. This positive, though unexpected, finding may be attributed to a number of causal factors. It may have arisen for users who are more familiar with the topic of the source paper, where the presented comparisons serves to overcome their problem fixation. It may be attributed to the (symmetric) visualizations that presented the source-to-target mapping or may be attributed to a number of other factors. Even if this specific situation triggers the need for further investigation, our system has a potential to identify such inspirations which could not be identified by human otherwise.

5.4. Inference Quality Evaluation

1000 inferences were generated and scored by the Additive Smoothing and Good-Turing methods. These scores were then used to categorize inference as High, Medium and Low, with the High category representing the best 20% of inferences, Low represents the bottom 20% and Medium are the remainder.

The top 20 inferences as scored by both techniques were collected, as were the weakest 20 inferences from both. Human ratings were then obtained for these inferences from 10 independent human raters, on a 5-point Likert scale (5 = Very good, 1 = very bad). Both methods showed a good ability to distinguish between good and bad inferences. The average score awarded to the High Category was Additive Smoothing (M=4.5) and Good-Turing (M=4.1), while for the Low category ratings were Additive Smoothing (M=2.2) Good-Turing (M=2.0). As can be seem s these techniques are more reliable at identifying Good inferences than bad ones. Overall, additive smoothing seems to offer the best potential at helping Dr Inventor at managing inference quality.

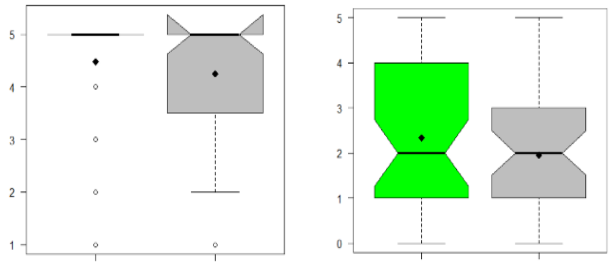

Figure 6. Scores awarded by Additive Smoothing (green) and Good-Turing (grey) to the inferences that people rated as good (left) and bad (right).

6. CONCLUSION AND FUTURE WORK

This paper described the Dr Inventor creativity support tool (CST) that aims to support scientific creativity by presenting novel analogical comparisons between publications. Firstly we presented the case for a CST based on the cognitive process of analogical thinking, describing how it might have a positive impact on the creativity of its scientist users.

We then described the major components of the Dr Inventor system. Dr Inventor is the first system to ever use "real" and automatically generated data from publications to simulate creative analogical thinking. It processes raw texts of scientific publications, generates graphs and analogically compares such graphs to identify analogies between documents. Based on the identified analogical similarity, Dr Inventor suggests inferences that can be transferred from the source for possible use in the target problem.

Thirdly, we presented an evaluation of the system to determine the level of creative support it provides to its users. We used the creative qualities of *novelty*, *usefulness* and *challenging the norms* to evaluate the level of inspiration and creativity support the system provides. The results indicated that Dr Inventor has a potential to identify novel and useful analogs. User ratings, of the analogies between pairs of papers identified by Dr Inventor, were provided by active researchers from computer graphics, using a 5 point Likert scale, with this feedback showing that the two highest rated comparison had many of the hallmarks of a good analogical comparison: high relational similarity, low conceptual similarity and a large number of inferences. The qualitative analysis indicates that Dr Inventor is capable of producing quality analogies and that these comparisons have a very beneficial impact on the creativity of the expert evaluators from the discipline of computer graphics.

Our future work will include co-references and causality to enhance the text analysis and in effect to improve the analogy mapping process. Another area of future work will focus on the metrics. Even if it is difficult to measure cognitive process, some preliminary results (relational and conceptual similarity) show that the correlation between users rating and the systems ranking could be improved by further enhancement of the metrics. Another future work that emerges from this research is the potential of creating a conceptual blend by merging analogical mappings of various papers.

ACKNOWLEDGEMENTS

The research leading to these results has received funding from the European Union Seventh Framework Programme ([FP7/2007-2013]) under grant agreement no 611383.

REFERENCES

[1] D. Gentner and L. Smith, "Analogical reasoning. In V. S. Ramachandran (Ed.)," in *Encyclopedia of Human Behavior (2nd Ed.).* , Oxford, UK, Elsevier., 2012, pp. 130-136.

[2] T. Brown, Making Truth: Metaphor in Science, University of Illinois Press, 2003.

[3] B. Shneiderman, "Creativity support tools," *Communications of the ACM,* vol. 45, no. 10, 2002.

[4] D. O'Donoghue, Y. Abgaz, D. Hurley, F. Ronzano and H. Saggion, "Stimulating and Simulating Creativity with Dr Inventor," in *Proceedings of the Sixth International Conference on Computational Creativity (ICCC 2015)*, Park City, Utah, 2015.

[5] D. Gentner and K. D. Forbus, "Computational Models of Analogy," *Wiley Interdisciplinary Reviews: Cognitive Science,* pp. 2(3):266-276, 2011.

[6] A. Koestler, The act of creation, 1964.

[7] O. Fuhrman and L. Boroditsky, "Cross-Cultural Differences in Mental Representations of Time," *Cognitive Science,* vol. 34, no. 8, pp. 1430-1451, 2010.

[8] T. S. Kuhn, The structure of scientific revolutions, 1962.

[9] K. Dunbar and I. Blanchette, "The in vivo/in vitro approach to cognition: The case of analogy," *Trends in cognitive sciences,* vol. 5, no. 8, pp. 334-339, 2001.

[10] M. Vendetti, A. Wu and K. Holyoak, "Far-out thinking generating solutions to distant analogies promotes relational thinking," *Psychological science,* vol. 25, no. 4, pp. 928-933., 2014.

[11] M. Boden, The Creative Mind, 2004.

[12] D. O'Donoghue and M. Keane, "A Creative Analogy Machine: Results and Challenges," in *4th International Conference on Computational Creativity*, Dublin, Ireland, 2012.

[13] F. Ronzano and H. Saggion, "Dr. Inventor Framework: Extracting Structured Information from Scientific Publications," in *Discovery Science*, Springer, 2015, pp. 209-220.

[14] B. C. Storm and G. Angello, "Overcoming fixation creative problem solving and retrieval-induced forgetting," *Psychological Science ,* 2010.

[15] M. T. Keane, T. Ledgeway and S. Duff, "Constraints on analogical mapping: A comparison of three models," *Cognitive Science,* vol. Cognitive Science, no. 18, pp. 387-438, 1994.

[16] M. Gick and K. Holyoak, "Analogical problem solving," *Cognitive psychology,* vol. 12, no. 3, pp. 306-355, 1980.

[17] V. Chaudhri, S. Heymans, A. Overholtzer, A. Spaulding and M. Wessel, "Large-Scale Analogical Reasoning," in *Proceedings of the Twenty-Eighth {AAAI} Conference on Artificial Intelligence*, Que'bec City, Canada, 2014.

[18] B. Fisas, R. Francesco and S. Horacio, "On the Discoursive Structure of Computer Graphics Research Papers," in *The 9th Linguistic Annotation Workshop held in conjuncion with NAACL 2015.* , 2015.

[19] A. Jordanous, "A standardised procedure for evaluating creative systems: Computational creativity evaluation based on what it is to be creative," *Cognitive Computation,* vol. 4, no. 3, pp. 246-279, 2012.

[20] R. Francesco and S. Horacio, "Knowledge Extraction and Modeling from Scientific Publications," in *In the Proceedings of the Workshop "Semantics, Analytics, Visualisation: Enhancing Scholarly Data" co-located with the 25th International World Wide Web Conference*, Montreal, Canada, 2016.

[21] C. Hamish, T. Valentin, R. Angus and B. Kalina., "Getting more out of biomedical documents with GATE's full lifecycle open source text analytics," *PLoS Comput Biol,* vol. 9, no. 2, pp. 1-16, 2013.

[22] A. Constantin, P. Steve and V. Andrei, "PDFX: fully-automated PDF-to-XML conversion of scientific literature," in *Proceedings of the 2013 ACM symposium on Document engineering.* , 2013.

[23] B. Bohnet, "Very high accuracy and fast dependency parsing is not a contradiction," in *Proceedings of the 23rd International Conference on Computational Linguistics. Association for Computational Linguistics, 2010*, 2010.

[24] H. Saggion, "SUMMA: A robust and adaptable summarization tool," *Traitement Automatique des Langues,* vol. 49, no. 2, 2008.

[25] D. Gentner, "Structure-mapping: A theoretical framework for analogy," *Cognitive Science,* vol. 7, pp. 155-170, 1983.

[26] A. Ruiz-Iniesta and O. Corcho, "A review of ontologies for describing scholarly and scientific documents.," in *Proceedings of 4th Workshop on Semantic Publishing (SePublica 2014)*, Aachen, Germany, 2014.

[27] B. Agarwal, S. Poria, N. Mittal, A. Gelbukh and A. Hussain, "Concept-level sentiment analysis with dependency-based semantic parsing: a novel approach," *Cognitive Computation,* vol. 7, no. 4, pp. 487-499, 2015.

[28] K. J. Holyoak and P. Thagard, "Analogical mapping by constraint satisfaction," *Cognitive Science,* vol. 13, pp. 295-355, 1989.

[29] D. Gentner and A. B. Markman, "Defining structural similarity," *Journal of Cognitive Science,* vol. 6, pp. 1-20, 2005.

[30] L. Cordella, P. Foggia, C. Sansone and M. Vento, "A (sub)graph isomorphism algorithm for matching large graphs," in *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2004.

[31] D. Lin, "An Information-Theoretic Definition of Similarity," San Francisco, CA, USA, 1998.

[32] G. A. Miller, "WordNet: A Lexical Database for English.," *Communications of the ACM,* vol. 38, no. 11, pp. 39-41, 1995.

[33] D. O'Donoghue, J. Power, S. O'Briain, F. Dong, A. Mooney, D. Hurley, Y. Abgaz and C. Markham, "Can a Computationally Creative System Create Itself? Creative Artefacts and Creative Processes," in *nternational Conference on Computational Creativity (ICCC)*, Ljubljana, Slovenia, 2014.

[34] K. J. Holyoak, L. R. Novick and E. R. Melz, "Component processes in analogical transfer: Mapping, pattern completion, and adaptation," in *Analogical connections. Advances in connectionist and neural computation theory*, vol. 2, Westport, CT, US, Ablex Publishing, 1994, pp. 113-180.

[35] S. F. Chen and J. Goodman, "An Empirical Study of Smoothing Techniques for Language Modeling," in *Proceedings of the 34th Annual Meeting on Association for Computational Linguistics*, Santa Cruz, California, 1996.

[36] W. A. Gale, "Good-Turing smoothing without tears," *Journal of Quantitative Linguistics,* vol. 2, no. 3, pp. 217-237, 1995.

[37] P. Jaccard, "Distribution de la flore alpine dans le bassin des Dranses et dans quelques régions voisines," *Bulletin de la Société Vaudoise des Sciences Naturelles,* vol. 37, pp. 241-272, 1901.

[38] L. Meng, R. Huang and J. Gu, "A Review of Semantic Similarity Measures in WordNet," *International Journal of Hybrid Informatioin Technology,* vol. 6, no. 1, 2013.

[39] M. Ramscar and Y. Daniel, "Semantic grounding in models of analogy: an environmental approach," *Cognitive Science,* vol. 27, no. 1, pp. 41-71, 2003.

[40] T. K. Landauer, P. W. Foltz and D. Laham, "Introduction to Latent Semantic Analysis," *Discourse Processes,* vol. 25, pp. 259-284, 1998.