Computer-Aided Patent Analysis: finding invention peculiarities

Gaetano Cascini[1], Davide Russo[1], Manuel Zini[2]
1 University of Florence - Department of Mechanics and Industrial Technologies, Methods and Tools for Innovation Lab, Florence, Italy
{name.surname}@unifi.it
2 DrWolf srl, Florence, Italy
mlzini@drwolf.it

Abstract. The application of standard Information Extraction techniques to Patent Analysis has several limitations partially due to the difference existing between patents and web pages, which are the object of the biggest majority of information search. Indeed, while in other fields customized processing techniques have been developed, the number of studies fully dedicated to patent text mining is very limited and the tools available on the market still require a relevant human workload. This paper presents an algorithm to identify the peculiarities of an invention through an automatic functional analysis of the patent text; as a result a ranked list of components and functions is provided as well as a selection of meaningful paragraphs disclosing the details of the invention. An example related to laser irradiation devices for medical treatment clarifies its basic steps.

1 Introduction

Today's text mining research activities are mostly dedicated to web content mining and encompass resource discovery from the Web, document categorization and information extraction from Web pages. The latter aims at the identification of the most relevant portion of a document and typically is based on the analysis of anchortexts, i.e. the visible part of hyperlinks. The rationale is that the larger is the number of anchortext terms in a sentence, the more relevant the sentence is likely to be, since it is supposed that the relevant sentences in the destination page are related with the anchortext in the source page [1]. Nevertheless, it is clear that this kind of approach is not applicable to patent analysis: the typical link between these documents is the citation, in which the anchortext is just the patent reference number, thus it is not related to conceptual details of an invention.

Another typical strategy consists in building ontologies to map terms relationships in a specific field, as performed in [2]. Such an approach is highly time consuming and thus it is still not widely applied.

Moreover patent examiners are skeptical about the adoption of software instruments substituting traditional Boolean search engines and manual efforts to perform prior art analyses. Nevertheless, they consider a top level priority the development of means to reduce the number of document to browse [3]. It can be stated that with the same purpose of reducing human involvement, also the reduction of the amount of text to be read from each document still is an essential goal.

In the past the authors have developed algorithms and tools for patent analysis aimed at:

- translating the description of an invention into a conceptual functional map [4];
- identifying knowledge flows between different fields of application [5];
- investigating the properties of Small World Networks as a base for Computer-Based idea generation system [6].

Among the crucial issues that emerged during those studies, the most challenging one is the identification of the most relevant part of a patent, i.e. the paragraphs disclosing the invention peculiarities. In other words, a relevant research goal is the capability to identify selected excerpts of the patent description, connected each other through the functional map of the invention, so that an expert in the field can focus his/her attention just on few sentences, instead of reading the whole text.

In [7] regular expressions and the analysis of the detail level of the description were presented as a means to achieve such a goal.

In this paper the adoption of a *tf-idf* (term frequency-inverse document frequency) ranking approach, to be performed after building a specific Thesaurus, is proposed as a means to highlight the relevant details of an invention and their disclosing sentences. The second section of the paper summarizes the previous results obtained by the authors and a comparison with other text-mining approaches is reported. Then, the proposed algorithm is detailed in section 3 and an exemplary application related to tumor ablation devices is shown in section 4 to demonstrate the efficiency of the proposed approach. A final discussion and opportunities for further developments conclude the last chapter.

2 Patent mining, related art

Patent mining is the branch of text mining technologies dedicated to the extraction of relevant information from patents and to their categorization. Indeed, just a few specialized tools fully dedicated to patent analysis exist, while typically general purpose text mining applications are adopted in combination with traditional Boolean search engines.

Commercially available patent databases provide basic means for information retrieval and citations tracking, but patents searches are still time consuming and require big efforts to be accomplished. In facts, citation analyses are the most used techniques for identifying within a company's patent portfolio the small number of valuable, high-impact patents against the large number of patents of marginal importance [8]. It is believed that a statistical analysis of the rate of publication of patents pertaining to a certain field or assigned to a certain company, provides information about technology maturity and corporate technology strategies. Typically, the analysis is performed by counting in an online database the number of

patents issued annually in a set of calendar years [9]. Besides, it normally takes five or more years from publication before a patent begins to be cited to any great extent. In general, 70% of all patents are either never cited, or cited only once or twice, so that even five citations place a patent in the top few percent of cited patents [10].

Therefore, the analysis of the free textual description is assuming a greater relevance for getting major advantages from disclosed inventions.

Text Mining applications provide effective means for content searches in the textual fields of electronic documents databases, but also the most recent works are not tailored for patent analyses as [11] and too often require a deep expertise about how to gain major advantages from this technology. Some special features are available in the Invention Machine Goldfire platform [12], mainly related to the application of syntactic parsing capabilities: each sentence is translated into a SAO triad (Subject, Action, Object), in order to produce a classification of the concepts contained in a patent description. Nevertheless, as well as for more traditional keywords based tools, no systems are available on the market for capturing the role of a component in an invention and for grouping patents according to their peculiar functionalities apart from their fields of application.

More specifically the following features aimed at speeding-up patent analysts activity still lack on the market:

– identifying the architecture of the claimed invention, distinguishing the functional (semantic) role of each component;

– identifying invention peculiarities as a means for providing an automatic extraction of the core of the patent; (it is worth to mention that too often the patent abstract is very low informative);

– clustering technical solutions according to the way a function is accomplished apart from the field of application (therefore providing proper means for technology transfer);

– allowing easy and effective queries by means of a multi-language taxonomic knowledge base so that search results do not depend on patent language and/or the use of synonyms, hyperonyms, meronymes etc.

Among the other activities, the authors have addressed the first issue by developing a novel analysis approach, disclosed in [4]. The proposed algorithm is capable of performing the functional analysis of an invention automatically, by processing the description and the claims of the related patent.

The algorithm basically consists in: (i) identifying the components of the invention; (ii) classifying the identified components in terms of detail/abstraction level and their compositional relationships in terms of supersystem/subsystem links; (iii) identifying positional and functional interactions between the components both internal and external to the system.

The components identification is performed taking into account that all the components must be referenced univocally to be identified in the patent figures. The following step of the analysis process is dedicated to the search of descriptive locutions (i.e. sentences containing verbs like “to form”, “to constitute” etc.) and specification expressions (like “the gripper of the pivot arm”) in order to identify subsystem/supersystem relationships, hence defining a hierarchy of detail/abstraction levels. Finally, positional and functional interactions between the identified components are determined by filtering, from the list of SAOs provided by a

syntactic parser, the triads containing irrelevant verbs (i.e. verbs like “to refer”, “to show”, as well as any other verb not describing some function or action).

Fig. 1 shows an exemplary application of the algorithm: the list of components and their hierarchical structure is represented by a tree; their functional interaction are mapped through a directed graph. These outputs, obtained fully automatically, i.e. without any interaction with the user, can be further processed in order to identify the peculiarities of the invention, as described below.

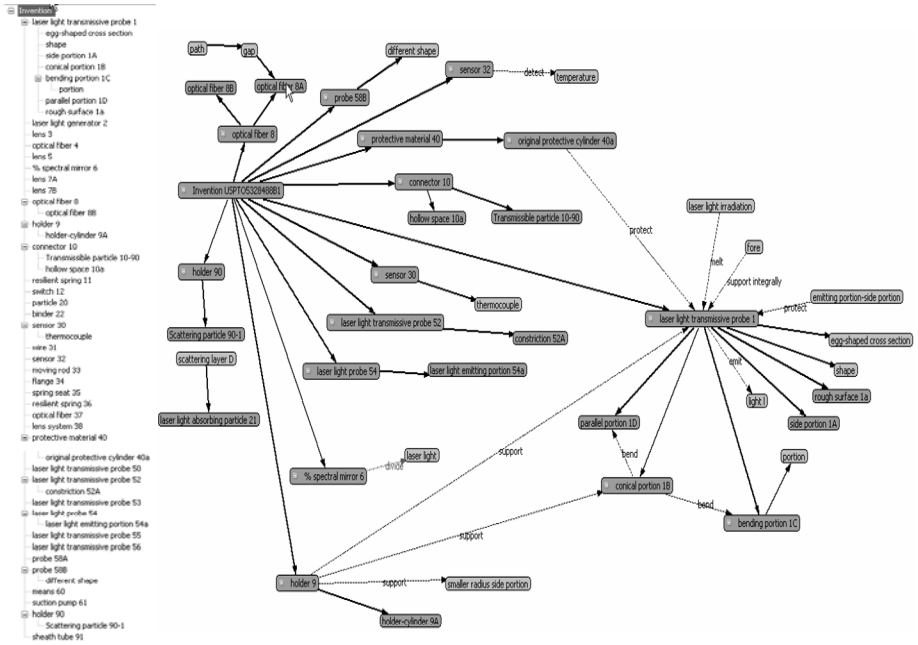


Fig. 1. Functional analysis of US patent 5,328,488 “Laser light irradiation apparatus for medical treatment”: list and hierarchy of components (left), functional map (right).

3 The functional subtraction algorithm

The rationale of the algorithm proposed in this paper is identifying the peculiarities of an invention by isolating the components and/or the functions of the system that differ from the state of the art. It is clear that the comparison among different patents should be independent from the language style of the inventor.

In order to fulfill this goal, the authors have developed an algorithm for building a Thesaurus of a specific set of patents.

It is worth to mention that during the components identification phase, alternative denominations can be found for each element of the system if several multi-words are referred to the same component reference number. For example in the patent US 6,161,390 – “Ice maker assembly in refrigerator and method for controlling the

same", the component 52 is called by the writer of the patent both "ice tray" and "ice container".

If two or more patents have a component sharing a common denomination, it is assumed that all their alternative denominations can be considered as synonyms forming a single Thesaurus entry. With this assumption the Thesaurus can be automatically built.

Such an assumption can be considered valid as far as the selected patents belong to a specific field of application, i.e. the same IPC class. Since the proposed algorithm for Thesaurus construction doesn't require any contribution of the user, it is clear that noisy terms may appear in the synonyms lists; nevertheless, according to the authors' experience, these noisy terms don't compromise the overall quality of the analysis. However, it is obvious that a manual validation of the extracted synonyms lists can contribute to a further improvement of the Thesaurus reliability.

Thus, when analyzing a set of patents related to a specific product/process, it is proposed to compare their components and functional interactions automatically extracted by means of the technique summarized in the previous section, taking into account the synonym lists gathered in the Thesaurus. The relevance of a component or a function *X* in a patent *k* is estimated according to the following formula, derived from the well known *tf-idf* weighting criterion [13]:

$$Score(X\ in\ Patent\ k) = \frac{\#of\ \ occurrences\ of\ \ X\ \ in\ k}{\max \#of\ \ occurrences\ in\ k} \log\left(\frac{\#of\ \ patents\ in\ the\ set}{\#of\ \ patents\ containing\ X}\right) \quad (1)$$

As stated above, the count of the number of occurrences and the count of the patents containing a certain component/function is performed assuming the equivalence of the synonyms gathered in the Thesaurus.

More in details, the proposed algorithm consists in the following steps:

1. Gather a set of patents related to a specific product/process;
2. Perform an automatic functional analysis of each patent according to the algorithm mentioned in section 2;
3. Build the Thesaurus of the patent set by identifying correspondences among the alternative denominations of the components of each patent;
4. Evaluate the score of each component of each patent according to the formula (1); evaluate the score of each functional interaction, i.e. each triad component-verb-component, according to the formula (1);
5. Identify the components at the highest detail level, i.e. the deepest leafs of the hierarchical tree built at step 2;
6. Extract from each patent the excerpt containing the top ranked components/functions.

4 Exemplary application of the proposed algorithm

A brief case study is here reported in order to clarify the algorithm described in the previous section and to demonstrate its validity. Instead of proposing the analysis of a big number of inventions that makes the manual validation rather critical, the

selected example is related to a small set of selected patents: the test set is here constituted by six patents related to laser irradiation devices for medical treatment. The set has been chosen because these patents are very similar to each other, they share the same title and the same inventor and, most of all, they deal with small improvements of the same device, thus the identification of their peculiarity requires a careful analysis also by a reader skilled in the art.

The test set is constituted by six patents titled "Laser light irradiation apparatus [for medical treatment]", assignee S.L.T. (Surgical Laser Technologies), issued from March 2003 to March 2005 and belonging to the US Classes 606/16 and 606/17: US 5193526, US 5209748, US 5290280, US 5328488, US 5496307.

Table 1. Patent US 5,328,488 "Laser light irradiation apparatus for medical treatment": excerpt from the list of components and their alternative denominations. Due to space limitations just the components with multiple denominations have been kept.

Component – ID Number	**Alternative denominations**
Laser light transmissive probe 1	laser light transmissive probe; probe; right side laser light transmissive probe; opposite laser light transmissive probe; laser light penetrating probe; transmissive probe; light transmissive probe; penetrating probe
rough surface 1a	rough surface; notch
optical fiber 8	optical fiber; single optical fiber
holder 9	holder; pinching holder
holder-cylinder 9A	holder-cylinder; holder
particle 20	particle; laser light scattering particle; scattering particle
wire 31	wire; lead wire
spring seat 35	rod; spring seat
resilient spring 36	resilient spring; spring
protective material 40	metal protective material; protective material; material
original protective cylinder 40a	original protective cylinder; cylinder
laser light transmissive probe 53	laser light transmissive probe; probe
laser light probe 54	laser light probe; penetrating probe; probe
laser light emitting portion 54a	laser light emitting portion; flat emitting portion

The application of the functional analysis algorithm (step 2) leads to the extraction from each patent of the followings:

- List of components and their alternative denominations (Table 1);
- Hierarchical tree of the components (Fig. 1, left);
- Functional interactions graph (Fig. 1, right).

Due to space limitations it is not possible to report the outputs arising from the analysis of each patent; besides, when dealing with a high number of patents, it is not convenient to start reading the details extracted from the individual patents, but it is suggested to start with an overall survey according to the ranking determined through the steps 4, 5. The bottom ranked components, i.e. those shared by the majority of the patents under comparison, constitute the common core of the examined system. The Table 2 reports the common components of the test set and

their alternative denominations provided by the Thesaurus. It can be stated that the novelty of the selected patents doesn't reside in the introduction of those components, while it might happen that the novelty consists in a modification of one of them aimed at providing a special property/functionality. The latter case can be identified as explained below.

In order to understand the rationale of the *tf-idf* criterion here applied, it is useful to focus the attention on the second and third column of Table 3, representing the components with the highest idf score and the exclusive forms/multi-words extracted respectively. It can be verified that most of the top idf-scored components are closely linked to the novelty disclosed in those patents. In facts, the "ballon" in US5,193,526, the "nipple" in US5,209,748, the "clads" in US5,290,280, the "constrictor" in US5,328,488 and the "fluid outlet" in US5,496,307 constitute the invention peculiarity or, at least, are strictly related to it. Those components point directly to the core of the invention and a person skilled in the art about laser irradiation devices for medical treatments will immediately understand what the patent deals with.

Table 2. Common components shared by the majority of the patents of the test set.

Reference Component	**Alternative denominations**
Laser light	laser light generator, right side laser light, constriction, portion, plural optical fibers, expose core
Fiber optic	optical fibers, original optical fibers, single optical fiber, original optical fiber, core
Layer	gold plate layer, laser light reflective layer, reflective layer, reflection layer, surface layer, concave surface
Probe	transmissive probe, penetrate probe, rough surface, emitter, laser light emitter, cylindrical-shaped emitter
Thermocouple	lead wire
Wire	guide wire
Tube	flexible protection tube, protection tube, holder tube, synthetic resin holder tube, main tube, core ,support tube, conductive tube, hole
Sheath	flexible sheath, sheath tube
Holder	metal holder, sleeve-like connector, hollow space
Lens	impinge lens, lens system

Table 3. Extraction of invention peculiarities through combined criteria.

US patent	idf score	Exclusive forms, multiwords	Detail level
5193526	Balloon 11	Hole 2A Impinge face 1A	
5209748	Nipple 3	Body 6A Screw hole 6C	
5290280	Emitter 20 Clads 1A,B Handle 5 Clad material 10	Cylindrical-shaped emitter 20A Knife-shaped flat emitter 20B Hook-shaped flat emitter 20C Claw-shaped emitters 20D Sickle-shaped emitter 20E Grip handle 5C Impinge lens 3	Emitter 20
5328488	Binder 22 Pump 61 Switch 12 Constrictor 52	Pinching holder	Laser light penetrating probe 58A
5496307	Covering 24 Fluid outlet 24a Fastener 20	Metallic fastener	Optical fiber 8B

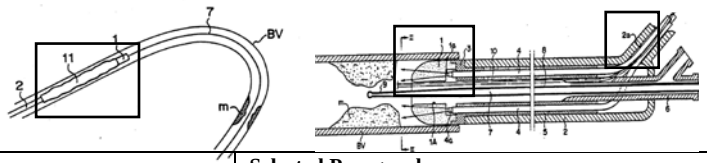
The exclusive forms/multi-words may draw the attention to properties and characteristics of invention details; for example, from US5,290,280 a number of characteristic shapes of the emitter are highlighted: cylindrical, flat knife, flat hook, claw, sickle. Further relevant features can be extracted by identifying the components at the highest detail level (step 5), i.e. the deepest leafs of the hierarchical tree built at step 2. The assumption here is that the description involves specific sub-components of the invention only if they are meaningful to the explanation of the invention itself.

The whole set of selected components and functions resulting from the steps 4 and 5 of the proposed algorithm can be used as seeds for a Content Analysis [14] of each patent The output is a selection of paragraphs where the top-ranked concepts are more represented (step 6). Again it can be stated that an expert in the field will be able to understand the content of those paragraphs without reading the whole document, at least for recognizing the relevance of the patent and its core novelty.

The top ranked paragraphs of the test set here adopted are reported in Table 4 and compared with the corresponding abstracts in order to demonstrate their higher informative content.

Table 4. Selected paragraphs of the test set.

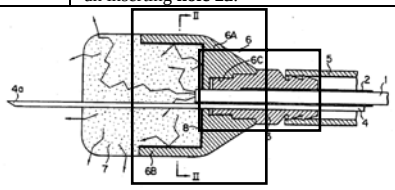
US 5,193,526



Patent Abstract	Selected Paragraphs
A laser light irradiation apparatus used for medical treatment of living tissues. According to a preferred embodiment, the apparatus comprises a probe and a plural number of optical fibers. The optical fibers surround the axis of the probe. Laser light goes through each optical fiber and is applied to the probe. Then, the laser light is emitted from the probe to	• By the laser light irradiation, the stricture part m is burnt off to widen the inside of a blood vessel. If desired, as shown in FIG. 4, pressurized air or pressurized liquid is sent into a **balloon 11** connected between a probe 1 and a main tube 2, thus, the **balloon 11** is expanded and press the stricture part m. As a result, together with the above mentioned burning off the inside of the blood vessel by the laser light irradiation, the

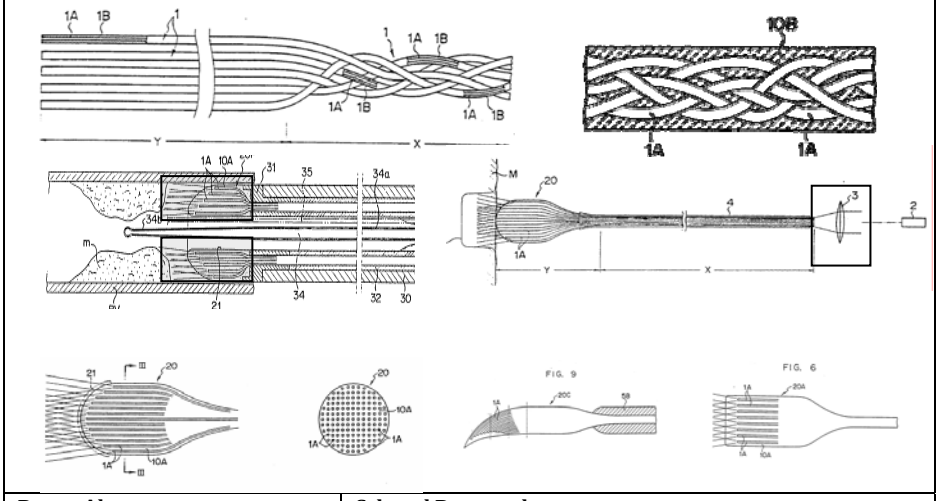
uniformly irradiate the tissues, and if desired, against the tissues over a broad area. Further, a guide wire and/or a lead wire for detecting a temperature can be placed so as to be coaxial with the probe. Therefore, a perforation of a normal part of the blood vessel can be prevented.	stricture part m can be broken mechanically. • Each tip portion of the optical fiber 1 is exposed to a core 4a. Each core 4a is adjacent to the back end face or the **impinging face 1a** of the probe 1. • Each optical fiber 4 is inserted into the main tube 2 from an inserting **hole 2a**.

US 5,209,748



Patent Abstract	Selected Paragraphs
A laser light irradiation apparatus used for medical treatment of tissues. According to a preferred embodiment, the apparatus comprises a probe, an optical fiber feeding laser light into the probe and a lead wire for detecting a temperature being inserted through and projecting from the probe. Then, the probe contains laser light scattering particles for uniform irradiation of the laser light against the tissues. Further, the probe is fabricated from a laser light tramissive synthetic material, and the fore end of a core of the optical fiber and the inserting part of the lead wire are in the synthetic material of the probe for easy molding for this apparatus.	• The fore end portion of the optical fiber 1 is inserted through a **nipple 3**, which is fabricated from a synthetic material such as polyethylene and the like. A lead wire 4 detecting a temperature having a thermocouple 4a at its fore end is provided alongside the optical fiber 1 and is also inserted through the **nipple 3**. • The holder 6 comprises a **body 6A**, which is tapered toward its back end, and a sleeve-like connector 6B, which has a hollow shape and which is projected from the **body 6A**. The screw of the **nipple 3** is adapted to mate with a connecting screw hole 6C of the holder 6 for connection. The optical fiber 1 and the lead wire 4 for detecting the temperature are inserted through the **body 6A**.

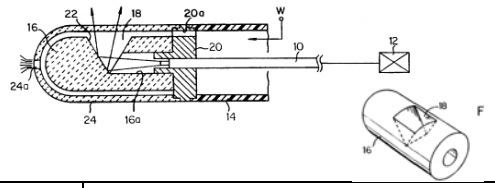
US 5,290,280



Patent Abstract	Selected Paragraphs
A laser light irradiation apparatus for medical treatment of living tissues, a preferred embodiment, comprises a laser light emitter and plurality of optical fibers. The fore end portion of each optical fiber is exposed to form an exposed light emitting core. The exposed cores are surrounded by a clad-material serving as the laser light emitter in order to reduce power loss of the laser light. Also, since there is no space between the emitting face of the optical fiber and the impinging face of the emitter, a cooling fluid is not required to pass through. The laser light is emitted from the emitter to	• Each original optical fiber 1 has a core 1A and a **clad 1B** surrounding the core 1A. Then, while at the base portion X, the fibers 1 are twisted in an irregular manner, the twisted original optical fibers 1 at the base portion X are heated to a temperature which is substantially the same temperature as or higher temperature than the melting point of the **clad 1B** and which is lower temperature than the melting point of the core 1A. Then, at the base portion X, the **clads 1B** of original optical fibers 1 are moulded to be one **clad 10B**, which includes the twisted cores 1A…. • As a result, a laser light **emitter 20** composing the resulting **clad-material 10A** and the number of cores 1A, which are arranged in parallel and which are included in the **clad-material 10A**. The shape of the laser light **emitter 20** corresponds to the shape of a container including the **clad-material 10A**. For example, as shown in FIG. 1, if the container has a constriction at the back end of the **emitter 20**, the shape of the laser light

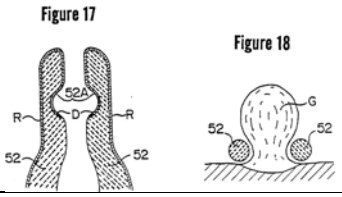
irradiate uniformly against the tissues, and if desired, against the tissues having a broad area. Further, a guide wire and a lead wire detecting a temperature can extend coaxially through the emitter. Therefore, a perforation of a normal part of the blood vessel can be prevented. To provide a more uniform power level distribution of the laser light, the optical fibers at the base portions are twisted.	**emitter 20** should be provided with an open having an inner diameter corresponding to the diameter of the constriction. • A **grip handle 5C** is provided at the back portion of the **emitter 20D** and can be operated with a restoring force. • The laser light irradiation apparatus of this type described above is used as follows. First, laser light fed from a laser light generator 2 goes through an **impinging lense 3**. • In the present invention, the **emitter** having several kinds of shapes can be applied. There are, for example, a **cylindrical-shaped emitter 20A** having a flat emitting face as shown in FIG. 6, a **knife-shaped flat emitter 20B** as shown in FIGS. 7 and 8, a **hook-shaped flat emitter 20C** as shown in FIG. 9, **claw-shaped emitters** 20D as shown in FIGS. 14, 15, 16 and 17, a **sickle-shaped emitter 20E** as shown in FIGS. 18, 19, 20 and 21.

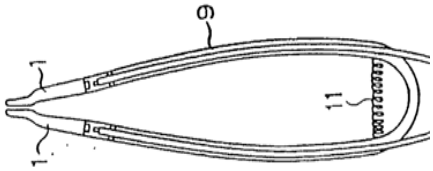
US 5,496,307



Patent Abstract	Selected Paragraphs
A laser light irradiation apparatus for medical treatment by irradiating an object tissue with laser lights transmitted through an optical fiber or fibers comprises a laser light reflector provided in front of the laser light emitting end of said optical fiber for reflecting the laser lights in a lateral direction of the apparatus, a covering which covers the reflector and is capable of transmitting the laser lights at least at the side portion thereof, wherein fluid is continuously supplied to a space between the covering and said reflector.	• A protection tube, separate from the covering, surrounding said optical fiber, said protection tube being linked with said **covering** by means of a **fastener** having a through hole, said **covering** being formed with a **fluid outlet** through which the supplied fluid having passed through said through hole and said protection tube can be discharged. • A **covering 24** which is made of, for example, light transparent ceramics such as heat resistant glass is linked with the front end of the protection tube 14 via the metallic **fastener 20** to enclose the reflector 16 therein. The **covering 24** is formed with a **fluid outlet 24a** at the front end thereof. Fluid, such as cooling water is supplied into a space between the protection tube 14 and the **optical fiber 10**. The flange of the **fastener 20** is formed with one or more through-holes 20a.

US 5328488





Patent Abstract	Selected Paragraphs
Laser light apparatus for medical treatment to permit amputations, incisions, vaporization of living tissues of an animal such as a human body, thermal therapy and the like. This apparatus consists of a laser light generator, a laser light transmissive probe system and a laser light transmitting system. The laser light transmissive probe system is provided with an opposed pair of laser light transmissive probes. The opposed pair of probes can be controlled by a medical operator so as to be moved into or out of contact with each other at their laser light emitting portions. Laser light is transmitted to the opposed pair of probes from the laser light generator through the laser light transmitting system. Then, a target area of living tissues is pinched by the opposed pair of laser light transmissive probes so as to be disposed between the opposed pair of laser light emitting portions.	• Further, the fore end portions of tho optical fibers 8A, 8B are provided in a U-shaped holder, in this embodiment a **pinching holder 9**. The **pinching holder 9** is made of metal and moves like a pincette. The above mentioned opposed pair of laser light transmissive probes 1, 1 are supported integrally by the fore end portions of the **pinching holder 9**. • On the other hand, as shown in FIGS. 17 and 18, when a projected tumor G formed on the surface of the tissues are excised, an opposed pair of laser light transmissive probes 52, 52 provided with pair of **constrictions 52A**, 52A can be used effectively. The opposed pair of laser light scattering layers D, D are preferably formed on the inner surface of the opposed pair of **constrictions 52A**, 52A respectively.

5 Conclusions and further developments

This paper presents an algorithm for patent analysis aimed at the identification of the invention peculiarities. Compared with standard Information Extraction techniques, the proposed approach is strongly based on typical patent features and first of all on the way the components of the invention are referred within the description. Such a characteristic allows to build a Thesaurus of the patent set under analysis that can be used to compare the inventions with reduced dependence from the language style of the author. Moreover, the comparison is not based on keywords directly extracted from the patent description, but relies on the identification of the invention components and their functional interactions.

The output of the proposed algorithm is a list of highlighted components and a small number of paragraphs representing the most relevant portion of the description; this excerpt of the patent is sufficient for an expert in the field to realize the scope of the invention and its core content. The algorithm has been applied to a small set of patents related to laser emitting devices for medical treatments to clarify the process and to demonstrate the advantages of the proposed approach.

The selected paragraphs can be used to improve the efficiency of clustering tools, both reducing the amount of text to be processed and removing the noisy part of the text, as proposed in [15].

A further improvement of the technique can be obtained by organizing the outputs according to a more comprehensive conceptual model, as proposed in [16]. Nevertheless, it is worth to note that in such a paper the authors identify the model elements by means of language patterns strongly dependent on the style of the writer.

References

1. Chen L., Chue W. L.: “Using Web structure and summarization techniques for Web content mining”, Information Processing and Management, vol. 41, 2005, pp. 1225–1242.
2. Krauthammera M., Nenadic G.: “Term identification in the biomedical literature”, Journal of Biomedical Informatics, Volume 37, Issue 6 , December 2004, Pages 512-526.
3. Krier M., Zaccà F.: “Automatic categorization applications at the European patent office”, World Patent Information, Volume 24, Issue 3 , September 2002, Pages 187-196.
4. Cascini G.: “System and Method for performing functional analyses making use of a plurality of inputs”, Patent Application 02425149.8, European Patent Office, 14.3.2002, International Publication Number WO 03/077154 A2 (18 September 2003).
5. Cascini G., Neri F., “Natural Language Processing for patents analysis and classification”, Proceedings of the TRIZ Future 4th World Conference, Florence, 3-5 November 2004, published by Firenze University Press, ISBN 88-8453-221-3.
6. Cascini G., Agili A., Zini M.: “Building a patents small-world network as a tool for Computer-Aided Innovation”, Proceedings of the 1st IFIP Working Conference on Computer Aided Innovation, Ulm Germany, November 14-15, 2005.
7. Cascini G., Russo D., “Computer-Aided analysis of patents and search for TRIZ contradictions”, International Journal of Product Development, Special Issue: Creativity and Innovation Employing TRIZ, Vol. 4, Nos. 1/2, 2007.
8. Breitzman A.F., Mogee M. E.: “The many applications of patent analysis”. Journal of Information Science, 28 (3), pp. 187–205, 2002.

9. Bigwood M.P.: “Patent Trend Analysis: Incorporate Current Year Data”. World Patent Information, Vol. 19, No. 4, pp. 243-249, 1997.
10. Karki M.M.S.: “Patent Citation Analysis: A Policy Analysis Tool”. World Patent Information, Vol. 19, No. 4, pp. 269-212, 1991.
11. Trappey A., Hsua F. C., Trappey C. V., Linc C. I.: “Development of a patent document classification and search platform using a back-propagation network”, Expert Systems with Applications, Vol. 31 (4), November 2006, pp. 755-765.
12. Goldfire Innovator, www.invention-machine.com.
13. Salton, G., McGill, M. J.: “Introduction to modern information retrieval”, McGraw-Hill, ISBN 0070544840, 1983.
14. Krippendorf K: “Content Analysis: An Introduction to Its Methodology”. Thousand Oaks, CA, Sage 2004.
15. Cascini G., Fantechi A., Spinicci E.: “Patent Clustering through selected functions”, submitted for publication to the Journal of Information Processing & Management, 2007.
16. Hui B., Yub E.: “Extracting conceptual relationships from specialized documents”, Data & Knowledge Engineering, Vol. 54 (1), July 2005, Pp. 29-55.