Thomson Innovation evaluation Patent content

Wytze Vlietstra

Date: 08-05-2016

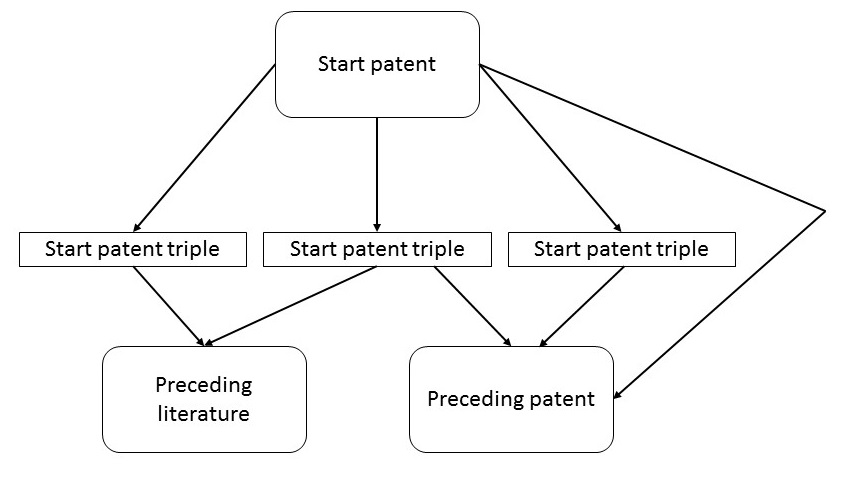
# Introduction:

For the ODEX project we are examining the additional data that Thomson-Reuters provides with their patents. This could be of use for the Limagrain use case, which wants to employ text mining (specifically relationship extraction) and structured data to compare claims between patents and literature.

# Short summary Limagrain use case:

The workflow is shown in the figure below, and works as follows:

1. Use a start patent as input
2. Collect text-mined triples (which should overlap with the claims) from this starting patent
3. Use the ODEX platform to collect preceding patents and/or literature that contain the same triples.



# Problem:

Text-mining patents is notoriously difficult, due to the complex legal formulations, and the often purposely obscure terminology used. The goal of the submitter of the patent is to make the contents and language in the patent as differing from preceding patents and literature as possible, to complicate comparison and enhance the patent’s chance of acceptance.

# Thomson-Reuters patent data:

Thomson-Reuters appear to offer a solution to the problem described above. They employ experts who re-write the contents of patents into a human-readable and more informative title and abstract, and add these to the patents. They attempt to use a standardized terminology (e.g. consistently use the term “maize”, even when the patent itself uses the term “corn”), but don’t appear to have an official terminology resource or a thesaurus/ontology. Thomson-Reuters also doesn’t extract triples from their patents. Perhaps we could provide these?

# Existing tools:

Thomson-Reuters themselves also provide text-analysis tools, with their Themescape and Clustering tools. These appear to be more based on term-vectors. While these are highly useful to obtain an overview of term-similarity and document-similarity of the input-documents, as of yet these tools are not capable of identifying novel, highly relevant documents based on these term-vectors, which would be an application more suitable for us.

# Content comparison example patent and Thomson-Reuters added data:

## Limagrain supplied patent

Limagrain has provided us with the following example patent: US 2015/0033406 A1. In this section we compare the contents of this patent with the additional fields provided by Thomson-Reuters (TR). Of note should be that the original patent is 174 pages long (of which 67 are the actual text), while the TR content is around 1-2 pages.

**Title:** *GENETIC MARKERS ASSOCIATED WITH DROUGHT TOLERANCE IN MAIZE*

**TR-Title:** *Making hybrid plant with enhanced water optimization involves crossing first plant containing specific genotype and second maize plant having specific genotype, and identifying at least one members of F1 generation*

**TR-Title terms:** *HYBRID PLANT ENHANCE WATER OPTIMUM CROSS FIRST CONTAIN SPECIFIC GENOTYPE SECOND MAIZE IDENTIFY ONE MEMBER GENERATE*

**Abstract:** *The presently disclosed subject matter relates to methods and compositions for identifying, selecting, and/or producing drought tolerant maize plants or germplasm. Maize plants or germplasm that have been identified, selected, and/or produced by any of the methods of the presently disclosed subject matter are also provided.*

**TR-Abstract: *Novelty:*** *Producing (P1) a hybrid plant with enhanced water optimization involves (a) providing a first plant containing a first genotype comprising any one of haplotypes A-M; (b) providing a second plant comprising a second genotype having any one of haplotypes A-M, where the second plant comprises at least one of haplotypes A-M that is not present in the first plant; (c) crossing the first plant and the second maize plant to produce F 1 generation; and (d) identifying at least one members of the F 1 generation that comprises a desired genotype including any combination of haplotypes A-M.*

***Use:*** *For producing a hybrid plant i.e. Zea mays plant with enhanced water optimization (claimed).*

***Advantage:*** *The method using enhanced water optimization provides increased or stabilized yield in water stressed environment as compared to a control plant; and can be planted at a higher crop density.*

## Second patent

In addition to the patent above, another patent was examined. This patent, US20150173318A1, was semi-randomly selected after searching for “drought maize”. This patent is 24 pages long,

**Title:** *NOVEL GENETIC FACTOR CAPABLE OF INCREASING YIELD IN MAIZE AND METHOD THEREOF*

**TR-Title:** *New maize plant, useful for developing further maize plants and hybrids with desired traits, e.g. additional leaves above the ear controlled by a genetic determinant which shows a non-dominant inheritance*

**TR-Title terms:** NEW MAIZE PLANT USEFUL DEVELOP HYBRID TRAIT ADD LEAF ABOVE EAR CONTROL GENETIC DETERMINE SHOW NON DOMINANT

**Abstract:** *A maize plant exhibiting additional leaves above the ear (LAE) architecture. The architecture is controlled by a genetic determinant, which shows a non-dominant inheritance. The maize plant is an inbred and hybrid variant with an enhanced yield. Also disclosed are methods for developing and producing the inbred and hybrid maize plant with an enhanced yield.*

**TR-Abstract: *Novelty***

*A maize plant exhibiting additional leaves above the ear (LAE) architecture, where the architecture is controlled by a genetic determinant which shows a non-dominant inheritance, is new.*

***Detailed Description***

*INDEPENDENT CLAIMS are:*

*seed of a maize plant above;*

*plant material obtainable from a maize plant above;*

*plant parts of a maize plant above;*

*maize kernels of a maize plant above;*

*pollen of the maize plant above;*

*an ovule of the maize plant above;*

*a tissue culture of regenerable cells of a maize plant above;*

*a maize plant regenerated from the tissue culture of (7), where the plant exhibit an additional LAE architecture, the architecture is controlled by a genetic determinant which shows a co-dominant inheritance;*

*a hybrid maize plant comprising a co-dominant genetic allele ELE1, where the allele is capable of conferring a phenotype of additional LAE; or a hybrid maize plant characterized by a genetic determinant conferring additional LAE phenotype, the genetic determinant being capable of transmission to progeny plants substantially as a single non-dominant gene;*

*maize seed derived from the hybrid maize plant of (9);*

*a method for producing seed maize with additional LAE architecture by pollinating a first maize plant with pollen of a second maize plant, where at least one of the maize plants possess a non dominant genetic determinant controlling additional LAE architecture, and harvesting seeds produced by the pollinated maize plant;*

*a method of producing a maize plant exhibiting an additional LAE architecture;*

*a method for producing hybrid seed maize exhibiting an additional LAE architecture by crossing first and second maize plants, where at least one of the maize plants is characterized by the presence of a genetic allele ELE1, conferring additional LAE phenotype, the genetic allele shows co dominant inheritance;*

*seed maize produced by the method of (11)-(13);*

*a maize plant produced by the method of (11)-(13);*

*hybrid maize seed, where the seed is characterized by a co-dominant ELE1 genetic allele conferring a phenotype with additional LAE in the maize plant grown from the hybrid maize seed;*

*a method for producing inbred maize seed characterized by the presence of a co-dominant ELE1 genetic allele by inbreeding a maize plant which is characterized by the genetic allele until the genetic composition of the progeny of such inbreeding becomes substantially stable;*

*maize seed produced by the method of (17);*

*a maize plant grown from the seed of (18);*

*a genetic determinant being inherited to progeny as a co dominant allele, where the genetic determinant is capable of conferring additional LAE architecture;*

*oligonucleotide sequences annealing with the genetic determinant of (20), where the sequences are suitable for the detection and production of maize plants having additional LAE architecture;*

*a method for increasing maize yield production to a commercially relevant extent in multiple geographical and/or whether-related environments or areas by growing in the geographical area maize plant;*

*a method of producing maize kernels or processed maize kernels as a food product by (a) providing a maize plant, (b) mutilating or propagating the maize plant, (c) allowing the plant to grow corn ears, and (d) harvesting the kernels of the corn ears; where the multiplication or propagation step (b) is performed by vegetative propagation; and*

*a maize field or maize greenhouse comprising plants above.*

***Activity***

*Plant Protectant. No biological data given.*

***Mechanism***

*None Given.*

***Use***

*The plant, seed and methods are useful for developing further maize plants and hybrids with desired traits, e.g. additional LAE architecture, increased yield, increased plant height, increased average kernel weight, increased number of kernels per ear, increased average kernel weight, increased average number of ears per plant, stress tolerance, improved prolificacy, parasite tolerance, pests tolerance, drought tolerance, earliness, adaptation to higher density planting, tolerance to striga, and improved and more efficient root system. The maize genetic determinants or elements, plants, seeds and plant products are used in multiple geographical and/or whether-related environments (e.g. tropical, sub-tropic, temperate, etc). The seed deposited under accession number NCEVIB 42074 is used for the production of maize kernels. The maize kernels are used as fresh produce, as fresh cut produce, or for processing such as, for example canning (all claimed).*

## Discussion

The TR-title and abstract appear to be more verbose and informative than the original title and abstract, although between our two examples the format does not seem to be completely consistent. Furthermore, the TR-title terms could be a helpful tool for the relationship extraction. There also appear to be less internal cross-references and “legalese” (e.g. “The method of claim 1, wherein said maize hybrid with enhanced water optimization…”). The European Patent Office solves this by creating a [claim-tree](http://worldwide.espacenet.com/publicationDetails/claims?DB=&ND=3&locale=en_EP&FT=D&date=20150129&CC=US&NR=2015033406A1&KC=A1&tree=true).

Furthermore, TR adds their own classification system to the data. While the original patent also contains a classification, I’m not aware about the differences between them.

A central advantage of using the Thomson-Reuters data is that it is translated from every major patent office in the world, diminishing the chance for false-negatives. However, as Limagrain has indicated they already have a fixed subset of patents they are interested in, it is questionable whether TR’s translated patents will be of use for the ODEX use case.

An additional benefit of the TR data is that it is already structured and cross-referenced. An XML format of patents should also be available from the various patent offices, but I’m not aware to which degree this is publically available (or whether a fee is required for these data).

# Thomson-Reuters:

Thomson-Reuters has expressed interest in using scientific publications as a means to gain publicity. Anneke and I have indicated we are willing to refer to TR in any relevant ODEX publications. Moreover, ODEX could be used to compare the quality of text-mined triples between the original patent and TR’s additions to the patent, potentially demonstrating a novel benefit and application of these data.

# Conclusions:

While not every feature is equally relevant for Limagrain’s ODEX use case, there appears to be a large potential benefit for including the Thomson-Reuters data in the text-mining process. However, these potential benefits cannot be guaranteed and remain to be investigated.