

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

CROWDSTRIKE, INC.,
Petitioner

v.

SKYSONG INNOVATIONS, LLC,
Patent Owner

Inter Partes Review Case No. IPR2025-01398
U.S. Patent No. 11,275,900

**PETITION FOR *INTER PARTES* REVIEW
OF U.S. PATENT NO. 11,275,900**

TABLE OF CONTENTS

I.	INTRODUCTION.....	1
II.	THE '900 PATENT.....	1
A.	THE '900 PATENT'S ALLEGED INVENTION	1
B.	THE '900 PATENT'S PROSECUTION	2
C.	A PERSON OF ORDINARY SKILL IN THE ART.....	3
III.	REQUIREMENTS UNDER 37 C.F.R. § 42.104	4
A.	STANDING UNDER 37 C.F.R. § 42.104(A)	4
B.	IDENTIFICATION OF CHALLENGE UNDER 37 C.F.R. § 42.104(B)	4
C.	CLAIM CONSTRUCTION UNDER 37 C.F.R. § 42.104(B)(3).....	5
1.	<i>“wherein the processor is configured to remove all child tags of the plurality of tags corresponding to a parent tag from the prediction list of tags where a probability value associated with the child [parent] tag is below a second predetermined threshold value.</i>	<i>6</i>
2.	<i>“wherein the processor is configured to add all parent tags of the plurality of tags corresponding to a child tag if a probability value associated with the parent [child] tag is above the first predetermined threshold value.”</i>	<i>8</i>
IV.	PRIOR ART SUMMARIES.....	9
A.	DELIU	9
B.	PENDAR	11
C.	FRANK	12
D.	RAMÍREZ-CORONA.....	13
E.	CHAWLA	14
F.	XIAO	15
G.	UKRAINCZYK	15
H.	NUNES	16
V.	GROUND 1: CLAIMS 1, 2, 4, 5, 10, 12, AND 13 ARE OBVIOUS OVER <i>DELIU</i> AND <i>PENDAR</i>.....	17
A.	CLAIM 1	17
1.	<i>Claim 1(Pre): A computer-implemented system for improving classification of criminal activities arising from cyber environments, comprising:</i>	<i>17</i>
2.	<i>Claim 1(a): a processor configured to:</i>	<i>17</i>

3.	<i>Claim 1(a)(1): access a ground truth dataset generated from deep web forum information, the ground truth dataset defining a predetermined tag hierarchy including predetermined tag labels corresponding topics of the deep web forum information;</i>	<i>18</i>
4.	<i>Claim 1(a)(2): access data related to a discussion topic for classification from a deep web forum;</i>	<i>23</i>
5.	<i>Claim 1(a)(3): extract a set of features from the data by assigning word vectors or paragraph vectors to the data; and</i>	<i>24</i>
6.	<i>Claim 1(a)(4): apply a machine classifier to the set of features to generate a prediction list of tags for classifying the discussion topic, wherein the prediction list includes a prediction probability value for each tag of the plurality of tags; and</i>	<i>25</i>
7.	<i>Claim 1(a)(5): add all parent tags associated with a tag of the plurality of tags to the prediction list based on a comparison between the prediction probability value for the tag and a first predetermined threshold value.....</i>	<i>29</i>
B.	CLAIM 2 - THE COMPUTER-IMPLEMENTED SYSTEM OF CLAIM 1, WHEREIN A PREDETERMINED TAG HIERARCHY IS DIVIDED INTO PARENT TAGS AND CHILD TAGS, WHEREIN A TOPIC REPRESENTED BY A CHILD TAG IS REPRESENTATIVE OF A SUBSET OF A TOPIC REPRESENTED BY A PARENT TAG, WHEREIN EACH CHILD TAG CORRESPONDS TO A PARTICULAR PARENT TAG, AND EACH TAG IS ASSIGNED A CORRESPONDING PROBABILITY VALUE.....	32
C.	CLAIM 4 – THE COMPUTER-IMPLEMENTED SYSTEM OF CLAIM 1, WHEREIN THE PROCESSOR IS CONFIGURED TO ADD ALL PARENT TAGS OF THE PLURALITY OF TAGS CORRESPONDING TO A CHILD TAG IF A PROBABILITY VALUE ASSOCIATED WITH THE PARENT TAG IS ABOVE THE FIRST PREDETERMINED THRESHOLD VALUE.	34
D.	CLAIM 5 - THE COMPUTER-IMPLEMENTED SYSTEM OF CLAIM 1, WHEREIN THE SET OF FEATURES ARE EXTRACTED USING TEXTUAL FEATURE EXTRACTION TO IDENTIFY TAGS ASSOCIATED WITH TOPIC TITLES OF THE DATA.	35
E.	CLAIM 10 - THE COMPUTER-IMPLEMENTED SYSTEM OF CLAIM 1, WHEREIN THE SET OF FEATURES ARE EXTRACTED FROM THE DATA USING A NEURAL NETWORK, WHEREIN THE NEURAL NETWORK IS CONFIGURED TO LEARN A CONTINUOUS DISTRIBUTED VECTOR	

	REPRESENTATION FOR DOCUMENTS AND WORDS WITHIN THE DATA.....	37
F.	CLAIM 12	39
	1. <i>Claim 12(Pre): A method, comprising:</i>	39
	2. <i>Claim 12(a): configuring a processor for executing operations including:.....</i>	39
	3. <i>Claim 12(a)(1): accessing data associated with a deep web forum, the data defining a topic for classification;.....</i>	39
	4. <i>Claim 12(a)(2): extracting a set of features from the data as inputs for a machine classifier; and.....</i>	40
	5. <i>Claim 12(a)(3): apply a machine classifier to the set of features to generate a prediction list of tags for classifying the topic, wherein the prediction list includes a prediction probability value for each tag of the plurality of tags; and</i>	40
	6. <i>Claim 12(a)(4): adding all parent tags associated with a tag of the plurality of tags to the prediction list based on a comparison between the prediction probability value for the tag and a first predetermined threshold value.</i>	40
G.	CLAIM 13 - THE METHOD OF CLAIM 12, FURTHER COMPRISING: OUTPUTTING AN UPDATE TO THE PREDICTION LIST OF TAGS, A TOTAL NUMBER OF THE TAGS OF THE PREDICTION LIST OF TAGS BEING DIFFERENT FROM THE UPDATE TO THE PREDICTION LIST OF TAGS TO MAINTAIN A PREDETERMINED TAG HIERARCHY.	41
VI.	GROUND 2 AND 3: CLAIM 3 IS OBVIOUS OVER <i>DELIU</i>, <i>PENDAR</i>, AND <i>FRANK</i> OR ALTERNATIVELY, OVER <i>DELIU</i>, <i>PENDAR</i>, AND <i>RAMÍREZ-CORONA</i>.....	43
A.	CLAIM 3 - THE COMPUTER-IMPLEMENTED SYSTEM OF CLAIM 1, WHEREIN THE PROCESSOR IS CONFIGURED TO REMOVE ALL CHILD TAGS OF THE PLURALITY OF TAGS CORRESPONDING TO A PARENT TAG FROM THE PREDICTION LIST OF TAGS WHERE A PROBABILITY VALUE ASSOCIATED WITH THE CHILD TAG IS BELOW A SECOND PREDETERMINED THRESHOLD VALUE.	43
	1. <i>Ground 2 – Deliu, Pendar, Frank</i>	43
	2. <i>Ground 3 – Deliu, Pendar, Ramírez-Corona</i>	45
VII.	GROUND 4: CLAIMS 6 AND 14 ARE OBVIOUS OVER <i>DELIU</i>, <i>PENDAR</i>, AND <i>CHAWLA</i>.....	48

A.	CLAIM 6 - THE COMPUTER-IMPLEMENTED SYSTEM OF CLAIM 1, WHEREIN A SET OF SYNTHETIC DATA SAMPLES ARE CREATED FOR ADDITION TO THE GROUND TRUTH DATASET TO ADDRESS CLASS IMBALANCE BY CREATING A SET OF SYNTHETIC SAMPLE POINTS ALONG A LINE DEFINED BETWEEN A FIRST MINORITY FEATURE VECTOR AND A SECOND FEATURE VECTOR AND ADDING THE SET OF SYNTHETIC SAMPLE POINTS TO THE GROUND TRUTH DATASET.....	48
B.	CLAIM 14 - THE METHOD OF CLAIM 12, FURTHER COMPRISING GENERATING A PREDETERMINED TAG HIERARCHY BASED ON A GROUND TRUTH DATASET, AND LEVERAGING SEMI-SUPERVISED LEARNING TO ADDRESS CLASS IMBALANCE AND INCREASE ACCURACY OF THE MACHINE CLASSIFIER.	51
VIII.	GROUND 5: CLAIM 7 IS OBVIOUS OVER <i>DELIU, PENDAR,</i> AND <i>XIAO</i>	52
A.	CLAIM 7 - THE COMPUTER-IMPLEMENTED SYSTEM OF CLAIM 1, WHEREIN THE GROUND TRUTH DATASET IS SUPPLEMENTED BY, FOR EVERY SAMPLE OF THE GROUND TRUTH DATASET IN A MINORITY CLASS, ADDING TOP SIMILAR DOCUMENTS FROM THE DEEP WEB FORUM INFORMATION, AS IDENTIFIED USING AN ELASTIC SEARCH SIMILARITY SCORE, AND ADDING THEM TO THE GROUND TRUTH DATASET.....	52
IX.	GROUND 6: CLAIMS 8 AND 9 ARE OBVIOUS OVER <i>DELIU,</i> <i>PENDAR,</i> AND <i>UKRAINCZYK</i>	55
A.	CLAIM 8 - THE COMPUTER-IMPLEMENTED SYSTEM OF CLAIM 1, WHEREIN THE DATA IS PARSED AND SEGMENTED INTO TOKENS, AND PREDETERMINED CHARACTERS AND PREDETERMINED COMMON WORDS ARE REMOVED FROM THE DATA.	55
B.	CLAIM 9 - THE COMPUTER-IMPLEMENTED SYSTEM OF CLAIM 1, WHEREIN WORDS WITHIN THE DATA ARE REDUCED TO THEIR ROOT FORM USING STEMMING AND LEMMATIZATION PROCEDURES.....	58
X.	GROUND 7: CLAIM 11 IS OBVIOUS OVER <i>DELIU, PENDAR,</i> AND <i>NUNES</i>.....	61
A.	CLAIM 11 – THE COMPUTER-IMPLEMENTED SYSTEM OF CLAIM 1, FURTHER COMPRISING: A CRAWLER IN OPERABLE COMMUNICATION WITH THE PROCESSOR AND OPERABLE TO TRAVERSE A PLURALITY OF DEEP WEB WEBSITES TO RETRIEVE INTERNET DOCUMENTS FORMING AT LEAST A PORTION OF THE DEEP WEB FORUM INFORMATION.	61

XI.	CONCLUSION.....	64
XII.	MANDATORY NOTICES UNDER 37 C.F.R. § 42.8(A)(1)	65
A.	REAL PARTIES-IN-INTEREST	65
B.	RELATED MATTERS	65
C.	LEAD AND BACK-UP COUNSEL	65
D.	37 C.F.R. § 42.8(B)(4) – SERVICE INFORMATION	66

TABLE OF AUTHORITIES

Cases:

<i>Fitbit, Inc. v. Valencell, Inc.</i> , 964 F.3d 1112, 1119–20 (Fed. Cir. 2020).....	6, 8
<i>KSR Int’l Co. v. Teleflex Inc.</i> , 550 U.S. 398, 416 (2007).....	31, 45, 54
<i>Pavo Sol.’s LLC v. Kingston Tech. Co., Inc.</i> , 35 F.4th 1367, 1373 (Fed. Cir. 2022).....	6, 8
<i>Phillips v. AWH Corp.</i> , 415 F.3d 1303, 1312 (Fed. Cir. 2005).....	5
<i>Volkswagen Grp. of Am., Inc. v. Michigan Motor Techs. LLC</i> , IPR2020-00446, Paper 9 at 10–11 (PTAB Aug. 25, 2020)	6, 8

Statutes:

35 U.S.C. § 102(a)(1)	10, 12-15
35 U.S.C. § 103(a)(2)	11
35 U.S.C. § 103	4, 5

Regulations:

37 C.F.R. § 42.6(e)	74
37 C.F.R. §42.8.....	73
37 C.F.R. § 42.8(a)(1)	65
37 C.F.R. § 42.8(b)(1)	65
37 C.F.R. § 42.8(b)(3)	65
37 C.F.R. § 42.8(b)(4)	65, 66
37 C.F.R. §42.24.....	73
37 C.F.R. § 42.100(b).....	5

37 C.F.R. § 42.104.....	4
37 C.F.R. § 42.104(a).....	4
37 C.F.R. § 42.104(b).....	4
37 C.F.R. § 42.104(b)(3)	5
37 C.F.R. § 42.105.....	74
37 C.F.R. § 42.105(B)	74

I. INTRODUCTION

CrowdStrike, Inc. (“Petitioner”) respectfully requests inter partes review (“IPR”) of Claims 1-14 (“Challenged Claims”) of U.S. Patent No. 11,275,900 (“’900 Patent”).

II. THE ’900 PATENT

A. *The ’900 Patent’s Alleged Invention*

The ’900 Patent describes “a computer-implemented system for improving classification of data associated with the deep web or dark net[.]” *’900 Patent* (Ex. 1001), Abstract, Fig. 6. Dark web environments provide an enormous repository of criminal data that requires automatic classification. *Id.*, 1:17-62. The ’900 Patent, therefore, involves “automatically assigning one or more labels (tags), in a hierarchical structure, to discussion topics seen in d2web [dark web-deep web] forums.” *Id.*, 3:11-15. The ’900 Patent refers to classifying d2web forums as a “multiclass, multi-label classification problem.” *Id.*, 3:40-43.

The data to be classified is naturally hierarchical. *Id.*, 5:16-17, 5:63-65. Therefore, the ’900 Patent leverages tag hierarchy to ensure consistency in its data. *Id.*, 5:62-6:6:67. For example, “for a predicted child tag every parent tag should also be predicted.” *Id.*, 6:5-7. The ’900 Patent provides three approaches for ensuring hierarchical consistency: adding parent tags, removing child tags, and a combination adding parent and removing child tags. *Id.*, 6:24-27. In the example of adding parent

tags, “[i]f a child tag is predicted by the classifier and if the corresponding parent tags in the hierarchy are not predicted, then all the parent tags are added to the prediction list.” *Id.*, 6:2-35. For removing all child tags, “all predicted child tags are removed that do not have all the hierarchical parent tags.” *Id.*, 6:34-43. A combination of adding parent and removing child tags is shown in example algorithm 1 below:

Algorithm 1 Combination of Adding Parent Tags and Removing Child Tags (T)
Input: Prediction list of tags (T) for each document. Output: Prediction list of tags that respect the tag hierarchy for each document. 1: $\beta \leftarrow$ remove child threshold value 2: $\alpha \leftarrow$ add parent threshold value 3: for t:= 1 to T do 4: $p(t) \leftarrow$ prediction probability of tag t. 5: if $p(t) < \beta$ then remove all child tags of tag t from prediction list 6: for t:= 1 to T do 7: $p(t) \leftarrow$ prediction probability of tag t. 8: if $p(t) > \alpha$ then add all parent tags of tag t from prediction list return updated prediction list

Id., Algorithm 1 (6:45-57).

B. *The '900 Patent's Prosecution*

The '900 Patent claims priority to provisional application 62/668,878, filed May 9, 2018. For the purposes of the Petition only, Petitioner applies May 9, 2018, as the priority date for the Challenged Claims.

In a response to a first office action, Applicant narrowed the claims to require, in relevant part, “the prediction list includes a prediction probability value for each

tag of the plurality of tags” and “add all parent tags associated with a tag of the plurality of tags to the prediction list based on a comparison between the prediction probability value for the tag and a first predetermined threshold value.” *’900 File History*, 284. Applicant argued “[t]he step of adding **all** parent tags...aims to solve the problem...observed that for some tags returned by the classifier function, the corresponding parent tags [are] not returned” such as when “the tag is wrongly predicted (false positive) or when some of the correct tags (false negatives) are not predicted.” *Id.*, 291 (emphasis original). Applicant differentiated prior art that only “evaluates inclusion of a parent tag or a child tag on a **case-by-case basis**[.]” *Id.*, 291-292 (“the present system adds ALL parent tags associated with a child tag if the child tag has a high enough probability prediction”) (emphasis original). A notice of allowance subsequently issued.

As set forth below, these concepts were well-known in the art before the earliest priority date of the ’900 Patent.

C. *A Person of Ordinary Skill in the Art*

A POSITA at the time of the ’900 Patent would have had a bachelor’s degree in computer science, electrical or computer engineering, or the equivalent, and 2-3 years of work/research experience in computer security and networking. *Dec.*, ¶¶54-

58.¹ Additional education (e.g., MS degree) might substitute for less experience and vice versa. *Id.*

III. REQUIREMENTS UNDER 37 C.F.R. § 42.104

A. *Standing Under 37 C.F.R. § 42.104(a)*

Petitioner certifies the '900 Patent is eligible for IPR and Petitioner is not barred or estopped from requesting IPR challenging the claims of the '900 Patent. Petitioner is not the owner of the '900 Patent, has not filed a civil action challenging the validity of any claim of the '900 Patent, and this Petition is filed less than one year after Petitioner was served with a complaint alleging infringement of the '900 Patent.

B. *Identification of Challenge Under 37 C.F.R. § 42.104(b)*

Petitioner requests the Challenged Claims be found unpatentable on the following grounds.

Proposed Grounds of Unpatentability	Exhibits
Ground 1: Claims 1, 2, 4, 5, 10, 12, 13 are obvious under AIA 35 U.S.C. § 103 over <i>Deliu</i> in view of <i>Pendar</i> .	Ex. 1004; Ex. 1005
Ground 2: Claim 3 is obvious under AIA 35 U.S.C. § 103 over <i>Deliu</i> , <i>Pendar</i> , and <i>Frank</i> .	Ex. 1004; Ex. 1005; Ex. 1007
Ground 3: Claim 3 is obvious under AIA 35 U.S.C. § 103 over <i>Deliu</i> , <i>Pendar</i> , and <i>Ramírez-Corona</i> .	Ex. 1004; Ex. 1005; Ex. 1008

¹ All citations to “Dec.” are to paragraph numbers in Ex. 1003, Declaration of Dr. Erez Zadok.

<u>Ground 4:</u> Claims 6 and 14 are obvious under AIA 35 U.S.C. § 103 over <i>Deliu, Pendar, and Chawla</i> .	Ex. 1004; Ex. 1005; Ex. 1009
<u>Ground 5:</u> Claim 7 is obvious under AIA 35 U.S.C. § 103 over <i>Deliu, Pendar, and Xiao</i> .	Ex. 1004; Ex. 1005; Ex. 1010
<u>Ground 6:</u> Claims 8 and 9 are obvious under AIA 35 U.S.C. § 103 over <i>Deliu, Pendar, and Ukrainczyk</i> .	Ex. 1004; Ex. 1005; Ex. 1011
<u>Ground 7:</u> Claim 11 is obvious under AIA 35 U.S.C. § 103 over <i>Deliu, Pendar, and Nunes</i> .	Ex. 1004; Ex. 1005; Ex. 1012

C. Claim Construction Under 37 C.F.R. § 42.104(b)(3)

In this proceeding, claims are interpreted under the *Phillips* standard. 37 C.F.R. § 42.100(b); 83 Fed. Reg. 197 (Oct. 11, 2018); *Phillips v. AWH Corp.*, 415 F.3d 1303, 1312 (Fed. Cir. 2005) (*en banc*). Petitioner applies the ordinary and customary meaning of the claim terms as understood by a POSITA for all terms not expressly construed below.²

² Petitioner’s challenge is not a waiver of any argument in litigation that any claim term in the ’900 Patent is indefinite or invalid. Petitioner does not waive its right to seek to construe claim terms in a different proceeding.

1. ***“wherein the processor is configured to remove all child tags of the plurality of tags corresponding to a parent tag from the prediction list of tags where a probability value associated with the ~~child~~ [parent] tag is below a second predetermined threshold value.*”**

For the purposes of this proceeding only, Petitioner proposes the Board correct a typographical error in Claim 3. The PTAB may “correct obvious minor typographical and clerical errors in patents.” *Pavo Sol. ’s LLC v. Kingston Tech. Co., Inc.*, 35 F.4th 1367, 1373 (Fed. Cir. 2022) (correcting “pivoting *the case*” to be “pivoting *the cover*”); *Fitbit, Inc. v. Valencell, Inc.*, 964 F.3d 1112, 1119–20 (Fed. Cir. 2020) (suggesting the standard from *Novo* extends to corrections the Board may make in post-grant proceedings, and holding it was error to not correct a “conspicuous” and undisputed error). “Correction is appropriate only if (1) the correction is not subject to reasonable debate based on consideration of the claim language and the specification and (2) the prosecution history does not suggest a different interpretation of the claims.” *Pavo Sol. ’s, LLC*, 35 F.4th at 1373 (internal citations omitted); *Volkswagen Grp. of Am., Inc. v. Michigan Motor Techs. LLC*, IPR2020-00446, Paper 9 at 10–11 (PTAB Aug. 25, 2020) (collecting cases).

The language of the claim and the intrinsic record make clear there is a scrivener’s error in the claim. Claim 3 as written requires the processor be configured to remove all child tags corresponding to a parent tag by checking if the probability value of “the child tag” is below a first predetermined threshold. First, there is no

antecedent basis for “*the* child tag.” As drafted, it is impossible to determine which child tag of the “all child tags” is thresholded against its probability value. Second, the intrinsic record clarifies this is a typographical error. The specification provides “Algorithm 1” that describes the claimed process for comparing tags to thresholds. Algorithm 1 describes removing child tags “of **tag *t***” when the probability “of **tag *t***” falls below a threshold:

<p>Algorithm 1 Combination of Adding Parent Tags and Removing Child Tags (T)</p> <hr/> <p>Input: Prediction list of tags (T) for each document. Output: Prediction list of tags that respect the tag hierarchy for each document.</p> <p>1: $\beta \leftarrow$ remove child threshold value 2: $\alpha \leftarrow$ add parent threshold value 3: for t:= 1 to T do 4: $p(t) \leftarrow$ prediction probability of tag t. 5: if $p(t) < \beta$ then remove all child tags of tag t from prediction list 6: for t:= 1 to T do 7: $p(t) \leftarrow$ prediction probability of tag t. 8: if $p(t) > \alpha$ then add all parent tags of tag t from prediction list return updated prediction list</p> <hr/>

’900 Patent, 6:45-58 (Algorithm 1, annotated)³. Here, tag *t* is the parent of any child tags. Therefore, using Algorithm 1, all child tags of the **parent tag** are added if the probability of the **parent tag** falls below a threshold. Patent Owner clarified this understanding during prosecution, stating that “claims 3 and 4 have been amended to provide further clarification that all...child tags are removed if **the parent tag** falls below a second probability threshold.” *’900 File History*, 293. Thus, the

³ All emphasis added unless otherwise noted.

applicant's characterization of claim 3 during prosecution matches Algorithm 1 and clarifies that the “**parent**” tag's probability value is checked against the second threshold, not the “child” tag.

2. *“wherein the processor is configured to add all parent tags of the plurality of tags corresponding to a child tag if a probability value associated with the ~~parent~~ [child] tag is above the first predetermined threshold value.”*

For the purposes of this proceeding only, Petitioner likewise proposes the Board correct a typographical error in Claim 4. *Pavo Sol.'s LLC*, 35 F.4th 1367 at 1373; *Fitbit, Inc.*, 964 F.3d at 1119–20; *Volkswagen Grp. of Am., Inc.*, IPR2020-00446, Paper 9 at 10–11.

Like Claim 3, the language of the claim and intrinsic record show there is a typographical error in the claim. Claim 4 as phrased requires the processor be configured to add all parents tags corresponding to a child tag by checking if the probability value of “the parent tag” is above a first predetermined threshold. Like above, there is no antecedent basis for “*the* parent tag[.]” There is no way of determining which parent tag of the “all parent tags” is thresholded against its probability value. Similarly, the intrinsic record proves this is a typographical error. Algorithm 1 also describes the process for adding parent tags, requiring adding parent tags “of **tag *t***” when the probability “of **tag *t***” exceeds a threshold:

Algorithm 1 Combination of Adding Parent Tags and Removing
Child Tags (T)

Input: Prediction list of tags (T) for each document.
Output: Prediction list of tags that respect the tag hierarchy for
each document.
1: $\beta \leftarrow$ remove child threshold value
2: $\alpha \leftarrow$ add parent threshold value
3: for t:= 1 to T do
4: $p(t) \leftarrow$ prediction probability of tag t.
5: if $p(t) < \beta$ then remove all child tags of tag t from prediction list
6: for t:= 1 to T do
7: $p(t) \leftarrow$ prediction probability of tag t.
8: if $p(t) > \alpha$ then add all parent tags of tag t from prediction list
return updated prediction list

'900 Patent, 6:45-58 (Algorithm 1, annotated). In this scenario, tag t is the child of any parent tags. Therefore, using Algorithm 1, all parent tags of the **child tag** are added if the probability of the **child tag** exceeds a threshold. Patent Owner confirmed this during prosecution, stating that “claims 3 and 4 have been amended to provide further clarification that all parent tags are added if the **child tag** exceeds a first probability threshold[.]” ’900 *File History*, 293. Thus, the applicant’s characterization of claim 4 during prosecution likewise matches Algorithm 1 and clarifies that the “**child**” tag’s probability value is checked against the first threshold, not the “parent” tag.

IV. PRIOR ART SUMMARIES

A. *Deliu*

Deliu published in the NTNU Open institutional repository and was publicly available at least by July 18, 2017. *Deliu* (Ex. 1004); *see also Munford Dec.* (Ex.

1006), ¶¶6-39; *Dec.*, ¶¶81-83. Namely, *Deliu* was publicly available as of July 18, 2017, on NTNU Open. *Munford Dec.* (Ex. 1006), ¶¶12-39. Ms. June Munford located the NTNU open repository webpage for *Deliu*, analyzed the submission “Metadata,” and confirmed that metadata matched the submission information on title page of *Deliu*—June 1, 2017. *Id.* As of September 2016, NTU Open allowed interested users to search for references using a variety of search parameters. *Id.* This same search functionality is available today. *Id.* *Deliu* is available in the NTNU library; was available since at least July 18, 2017; and simple search queries including keywords from the title and abstract led to locating *Deliu*. *Id.* This methodology would have led an interested party to locate *Deliu*. *Dec.*, ¶¶81-84. *Deliu* is prior art under at least 35 U.S.C. § 102(a)(1).

Deliu compares the performance of machine learning approaches in “Extracting Cyber Threat Intelligence From Hacker Forums[.]” *Deliu*, 1. *Deliu* obtains a leaked dataset from a hacker forum called Nulled.IO. *Deliu*, 62. A labeled dataset is constructed by classifying the forum posts into irrelevant and relevant posts based on whether the posts contain one of a number of hand-selected keywords related to security. *Deliu*, 62-63. The relevant posts were then further classified into a number of information security categories. *Deliu*, 63, 74-76, Fig. 12.

Deliu teaches using well-known machine learning techniques to classify forum posts. *Deliu*, 7 (“In this thesis, we explore the capabilities of Machine

Learning methods in the task of locating relevant threat intelligence from hacker forums.”). For example, *Deliu* teaches vectorizing text using Bag-of-Words and n-gram feature extraction techniques, and feeding those feature vectors to k-NN, Decision Tree, and SVM classification algorithms. *Deliu*, 64-66. *Deliu* also teaches vectorizing text using word2vec, GloVe, and Random vector feature extraction algorithms for analysis by a Convolutional Neural Network. *Deliu*, 66-70.

B. *Pendar*

U.S. Patent No. 9,330,167 to Pendar (“*Pendar*”) (Ex. 1005) was filed on May 13, 2013, qualifying as prior art to the ’900 Patent under at least 35 U.S.C. § 102(a)(2).

Pendar teaches a system and apparatus for classifying and tagging textual data. *Pendar*, 7:58-61. (“Methods, apparatus and computer program products described herein are operable for the classification and tagging of textual data using automatically learned or generated queries”). Specifically, *Pendar*’s “classification and tagging apparatus 100, may be configured for accessing machine-readable text related to a service provider.” *Id.*, 15:33-35. After accessing the textual data, *Pendar*’s apparatus may identify “one or more services provided by the service provider[.]” *Id.*, 15:39-41. Based on this identification, *Pendar*’s apparatus may then assign “one or more labels to the service provider.” *Id.*, 15:46-48. Labels can be assigned by *Pendar*’s apparatus scoring a document and comparing that score to a

predetermined threshold. *See e.g., Pendar*, 14:37-57 (describing documents being scored, a comparison of document score to a label’s threshold value, and assignment of the label to the document if the score exceeds the threshold).

Pendar teaches that the labels may be related to one another based on a hierarchy. *Pendar*, 15:52-53 (“In another embodiment, one or more labels may be determined to be related to another label in the hierarchy”). For example, *Pendar* discusses that the label “‘burgers’ may be associated to ‘traditional American cuisine’ and ‘meat’,” so the hierarchical structure of the service taxonomy “includes at least [three] levels.” *Id.*, 15:53-57. Further, as the labels may be related to one another in a hierarchy, *Pendar*’s apparatus may be configured to assign “a label for each dominant category of each of the one or more assigned service labels.” *Id.*, 15:63-66, Fig. 6.

C. *Frank*

Frank published in the University of Waikato Library and was made publicly available in 2001. *Frank* (Ex. 1007); *see also Munford Dec.* (Ex. 1006), ¶¶57-62; *Dec.*, ¶¶183-186. Namely, *Frank* was publicly available as of January 8, 2001, at least based on its acquisition and publication at the University of Waikato Library. *Munford Dec.* (Ex. 1006), ¶¶57-62. Ms. June Munford located the University of Waikato Library thesis collection, located *Frank* and its MARC record, and confirmed the MARC record explains that *Frank* was made available as of January

8, 2001. *Id.* Simple search queries including keywords from the title, abstract, and/or author name led to locating *Frank*. *Id.* This methodology would have led an interested party to locate *Frank*. *Dec.*, ¶¶183-186. *Frank* is prior art under at least 35 U.S.C. § 102(a)(1).

Frank teaches “pruning algorithms for decision trees and lists that are based on significance tests.” *Frank*, 3. *Frank* discusses many pruning methods, including “Critical Value Pruning[.]” *Id.*, 76. “Critical value pruning” compares the “splitting criterion” value of a node to a threshold value to make pruning decisions. *Id.* Specifically, “[w]hen a subtree is considered for pruning, the value of the splitting criterion at the corresponding node is compared to a fixed threshold, and the tree is replaced by a leaf if the value is too small.” *Id.*

D. *Ramírez-Corona*

Ramírez-Corona published in Probabilistic Graphical Models (PGM 2014) and was made available in 2014. *Ramírez-Corona* (Ex. 1008); *see also* *Munford Dec.* (Ex. 1006), ¶¶48-51. *Ramírez-Corona* is prior art under at least 35 U.S.C. § 102(a)(1).

Ramírez-Corona teaches a hierarchical multi-label classification (HMC) method “for tree and Directed Acyclic Graphs (DAG) hierarchies.” *Ramírez-Corona*, 1. Specifically, *Ramírez-Corona* compares pruning approaches for Non Mandatory Leaf Node Prediction (NMLNP) which varies the pruning direction,

time, and condition. *Id.* *Ramírez-Corona* addresses two different directions for pruning, “Top-Down” and “Bottom-Up,” as well as different times to perform the pruning stage (i.e., “pruning time”) *Ramírez-Corona*, 7. Further, three pruning conditions, which “establish[] the condition to fulfill to prune the node,” are used in *Ramírez-Corona*’s comparison. *Id.*

E. *Chawla*

Chawla published in the Journal of Artificial Intelligence Research and was made available in June 2002. *Chawla* (Ex. 1009); *see also* *Munford Dec.* (Ex. 1006), ¶¶40-43. *Chawla* is prior art under at least 35 U.S.C. § 102(a)(1).

Chawla teaches an approach to construct “classifiers from imbalanced datasets[.]” *Chawla*, 1. An imbalanced dataset is one in which the “classification categories are not approximately equally represented.” *Id.* *Chawla*’s proposed approach is an “over-sampling approach” that oversamples the minority class by generating “synthetic” samples (referred to as “SMOTE”). *Id.*, 6-8.

Chawla teaches that the first step to generate synthetic samples is to “[t]ake the difference between the feature vector (sample) under consideration and its nearest neighbor.” *Chawla*, 8. Then, the difference is multiplied “by a random number between 0 and 1” and added to the feature vector. *Id.* A random point along the line segment between the two features is then selected, and ultimately, this

process forces “the decision region of the minority class to become more general.”

Id.

F. *Xiao*

Xiao published in Proceedings of NTCIR-7 Workshop Meeting and was made available in December 2008. *Xiao* (Ex. 1010); *see also Munford Dec.* (Ex. 1006), ¶¶52-56. *Xiao* is prior art under at least 35 U.S.C. § 102(a)(1).

Xiao teaches a system that automatically maps research papers, based on their abstracts, “into patent classification taxonomy, namely International Patent Classification (IPC).” *Xiao*, 1. In this system, *Xiao* utilizes a KNN-classifier that calculates “the similarity between an input document and each document in the training data set.” *Id.*, 3. In particular, *Xiao* discusses using five methods to calculate this similarity, and includes the “BM25” weighting method as one of the similarity calculation methods. *Id.*

G. *Ukrainczyk*

U.S. Patent No. 7,028,250 to Ukrainczyk (“*Ukrainczyk*”) (Ex. 1011) was filed on May 25, 2001, qualifying as prior art to the ’900 Patent under at least 35 U.S.C. § 102(a)(2).

Ukrainczyk teaches “a method for simultaneously classifying at least one document into a plurality of categories is disclosed.” *Ukrainczyk*, 3:27-29. The document being classified is “parsed into a plurality of unique tokens” to create a

document feature. *Id.*, Abstract; 3:35-37. A feature vector is created for the document, which includes a count corresponding to each document feature and is ultimately multiplied by a category vector to determine if the document should be classified into a particular category. *Id.*, 3:37-50.

H. *Nunes*

Nunes published in Proceedings of the 2016 IEEE Conference on Intelligence and Security Informatics and was available to the public at least by November 8, 2016. *Nunes* (Ex. 1012); *see also Munford Dec*, ¶¶44-47. *Nunes* is prior art under at least 35 U.S.C. § 102(a)(1).

Nunes teaches a “system for cyber threat intelligence gathering from various social platforms on the Internet particularly sites on the darknet.” *Nunes*, Abstract. *Nunes*’ system includes a “crawler” that is “a program designed to traverse the website and retrieve HTML documents.” *Id.*, 2. The crawlers used by *Nunes* are specifically designed for different forums. *Id.*, (“We have designed separate crawlers for different platforms (markets/forums) identified by experts due to the structural difference and access control measures for each platform”).

V. GROUND 1: CLAIMS 1, 2, 4, 5, 10, 12, AND 13 ARE OBVIOUS OVER DELIU AND PENDAR

A. Claim 1

1. Claim 1(Pre): A computer-implemented system for improving classification of criminal activities arising from cyber environments, comprising:

To the extent the preamble is limiting, *Deliu* teaches a computer-implemented system that compares “the classification performance of different machine learning algorithms on data from hacker forums.” *Deliu*, 65-66. *Deliu*’s machine learning algorithms classify textual data from hacker forums, used by cyber criminals to exchange cyber attack information. *Id.*, 18 (“Different communication channels such as online forums are being used by cyber criminals to exchange these services”); *see also id.*, 17 (describing analyzing “Dark Web” forum posts from “[c]yber criminals (also called hackers)”). *Deliu*’s machine learning techniques “improve” upon manual analysis, automating or semi-automating the analysis “to enrich [] cyber security controls.” *Id.*, 18-19.

2. Claim 1(a): a processor configured to:

Deliu teaches “[t]he computer system used for running the experiments” uses an “Intel Quad Core i7-3820 3.60GHz” processor. *Deliu*, 73. *Deliu*’s processor is “configured to” execute its experiments via tools and software, including Python 3.6.0, Sklearn 0.18.1, Theano 0.8.2, and MySQL Community Server (GPL) 5.7.17-log. *Id.*, 73.

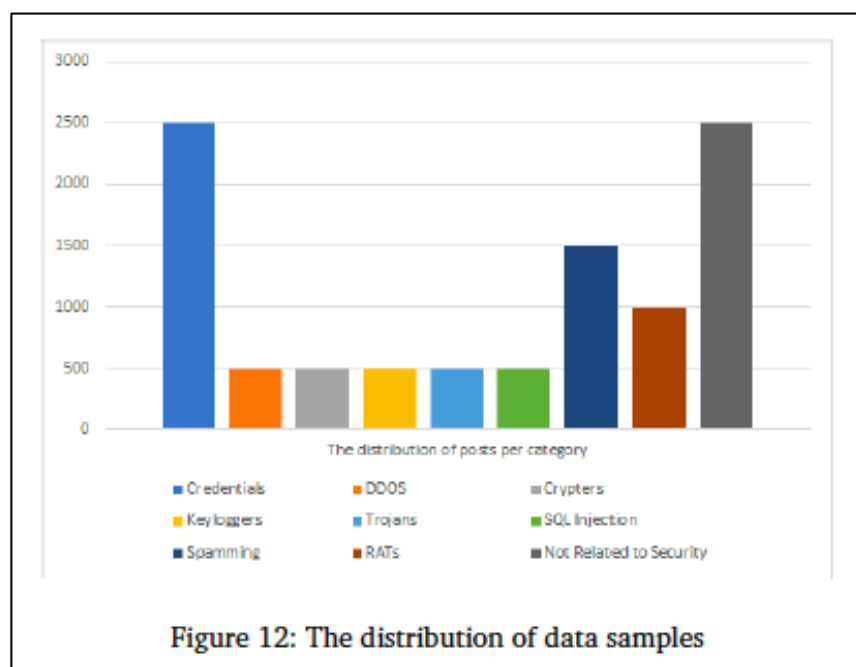
3. ***Claim 1(a)(1): access a ground truth dataset generated from deep web forum information, the ground truth dataset defining a predetermined tag hierarchy including predetermined tag labels corresponding topics of the deep web forum information;***

Deliu teaches accessing a multi-class dataset from the Nulled.IO hacker forum (i.e., *ground truth dataset generated from deep web forum information*) comprising labeled posts (i.e., *including predetermined tag labels*). *Deliu*, 62-64. These labels correspond to topics in the hacker forum (i.e., *corresponding [to] topics of the deep web forum information*”), such as DDOS, Trojans, SQL Injections, etc. *Id.*; *see also id.*, 76 (Fig. 12).

In more detail, *Deliu* tested its machine learning classification methods “on a **real hacker forum.**” *Id.*, 7, 19, 20; *see also Id.*, Title (“Extracting Cyber Threat Intelligence From Hacker Forums”). Namely, *Deliu* tests its experiments on a leaked dataset generated from the Nulled.IO dark-web forum. *Id.*, 62, 73-74. While *Deliu* obtained the forum information from a whistleblower website (<http://leakforums.net/thread-719337>), a POSITA would have understood Nulled.IO is a deep web forum. *Dec.*, ¶¶96-97. The Nulled.IO forum was originally part of the dark web. *Deliu*, 17-18 (“In this thesis, we focus on the last source, Dark Web...we study hacker forums as they represent a considerable part of Dark Web”); *Dec.*, ¶96. The dark web is a subset of the deep web. *Id.*, ¶¶96-97; *'900 Patent*, 1:63-64 (“The dark web forms a small part of the deep web that is not indexed by web search

engines.”). Therefore, *Deliu*’s Nulled.IO dataset is “generated from deep web forum information.” *Dec.*, ¶¶96-97.

Like the ’900 Patent, *Deliu* manually labels its ground truth dataset with “predetermined tag labels[.]” *Compare* ’900 Patent, 3:61-62 (“The ground truth to build a machine learning model **may be hand labeled by field experts**”) *with Deliu*, 62 (“In order to build a model which is able to filter out posts which are not relevant to security, the presence of labeled samples is required...we constructed a dataset by manually labeling some of the posts in the identified forum”). *Deliu* teaches “construct[ing] a multi-class dataset” from the Nulled.IO forum and manually labeling posts with a variety of predetermined tag labels. *Deliu*, 74-76. The distribution of labeled samples per category is shown below in Figure 12:



Id., 76 (Fig. 12).

These labels *correspond to topics of the deep web forum information*. Table 7 below provides illustrative dark web forum posts for each of the categories in the “multi-class dataset”:

Table 7: Examples posts from multi-class dataset	
Category	Example
Not Related to Security	Hello dear nullified.io community. This is a simple question, what are your favourite movies? ; ; Mines? Idk. Probably Jackie chan movies and/or taken series
Credentials	Hello Guys, just cracked these spotify accounts. It's premium. ; Enjoy. ; ; [hide] seand.bertran@gmail.com:Daniel81588 niru660@gmail.com:airbusa380 pskinner63@gmail.com:este11a ellis.nathan@gmail.com:unit24 jano761012@gmail.com:324657 sammy.pierce@gmail.com:kre8tv12 zelus.et.radix@gmail.com:ocelote csoto1251@gmail.com:blah123123 jcar-rara24@gmail.com:maryjane yourwifealex@gmail.com:tita2165 ; [/hide] ;
DDOS attack	If i want to ddos someone, which Port should i use? and which method is best? UDP? ; 80 UDP is ok .
Keyloggers	Source code of a simple, little Keylogger for Windows developed in C++ for Windows only. A good starting point to develop a private FUD keylogger. ; ; Reply and upvote to unlock the source code link: it
Crypters	Hey guys! ; Here is the cracked Codelux version 3.6.6 It is a crypter to make files undetected by AV. Deobfuscated by li0nsar3c00l, cracked by Meth ; Here is the product description : ; ; Download ; [hide]https://mega.co.nz/#!K1wjJLDJ!YuTqbnlPVhjo6ivT-oCKX6G1G3CdyLfE3FuYmua1J6w[/hide]
Trojans	Enjoy ; ; http://www12.zippyshare.com/v/5AQ1WVsB/file.html TROJAN DETECTED!
SQL Injection	20 SQL injectable's. ; [hide] http://allaces.ru/p/episode.php?id=1' http://www.keswick2barrow.co.uk/faqs.asp?ID=1' http://www.alicantegolf.org/principal.php?idp=1' http://www.bulsu.edu.ph/news.php?id=999999.9 union all select 1,2,3,4,5,6,7777777,8,9,10 http://www.wollin.de/w3.php?nodeId=3'3 http://www.abramat.org.br/site/lista.php?secao=999999.9 union all select 1 ; ; [/hide] ; Have fun!
Spamming	User has 4 posts and spammed VIP section with 3 different fake scripts. Asking for a ban.
RATs	I think, personally, that this RAT is one of the best for its compatibility with all windows machines. ; ; Download Here: [hide] https://megabitload.com/download/index/96758211/ [/hide]

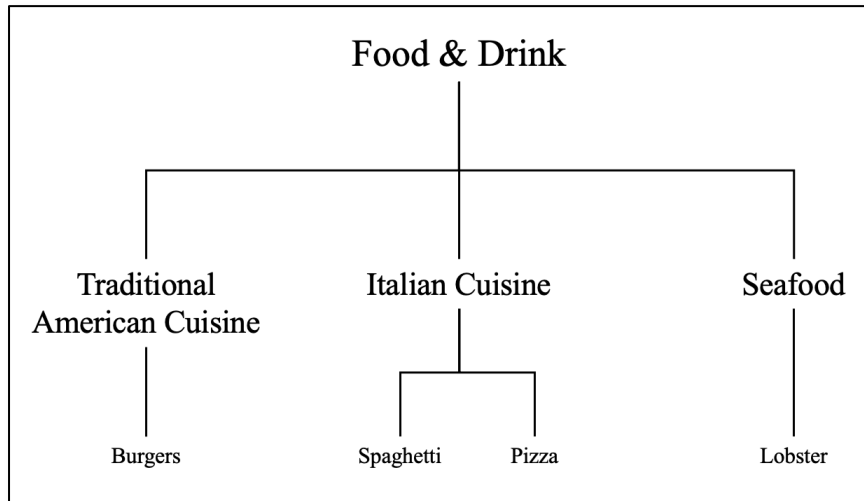
Id., 77 (Table 7).

Deliu “accesses” the multi-class dataset to train and test its classifier algorithms. *Id.*, 80-86. *Deliu* “measure[s] the performance of traditional classifiers...by averaging the results of running 10-fold cross validation for 10

times” such that in each iteration, “a random sample of approximately 10% is hidden from the training algorithms and used as test data.” *Id.*, 80 (describing process for the binary dataset). This same process is used for the multi-class dataset. *Id.*, 85 (“We run all of the experiments reported so far in the multi-class dataset as well”).

Deliu does not specifically discuss that its multi-class dataset includes hierarchical data such that it defines a “predetermined tag ***hierarchy***[.]” *Pendar* teaches this.

Pendar teaches a corpus-database of prelabeled documents where the labels are taken from a defined hierarchy. *Pendar*, 7:61-64 (“a training method...may be configured to access a corpus that is comprised of annotated, labeled, tagged, or otherwise classified documents or textual portions to generate a set of features”). The “corpus may comprise one or more documents with an associated label such as a category (e.g., food and drink) and/or service (e.g., pizza) taken from **a defined service hierarchy or service taxonomy**.” *Id.*, 8:30-36. The service taxonomy provides a “**predefined hierarchical structure** that includes at least [three] levels.” *Id.*, 15:55-57. In an example, the label “burgers” is associated with “traditional American cuisine” which is associated with “food and drink[.]” *Id.*, 15:46-62. *Pendar*’s service taxonomy is depicted in the below annotation:



Dec., ¶¶103-105 (Illustration based on *Pendar*, 8:32-36, 15:1-23, 15:46-62).

A POSITA would have been motivated to hierarchically label *Deliu*'s multi-class dataset pursuant to *Pendar* to improve *Deliu*'s classification techniques. *Dec.*, ¶¶107-113. *Deliu* expressly envisions utilizing other deep learning algorithms to compare to the algorithms tested in its experiments. *Deliu*, 108 ("Future Work"). One such improved deep learning algorithm is a hierarchical SVM. *Dec.*, ¶¶107-109. Hierarchical SVMs were known in the art and were known to provide improved data classification. *Dec.*, ¶¶108-110; Ex. 1015 (*Hoffman*), 9 ("hierarchical SVM outperforms flat SVM"). A POSITA would have recognized that by implementing a hierarchy or taxonomy among the labels, the classification performance of the Support Vector Machine would be more accurate and representative of real-world classification systems. *Dec.*, ¶¶108-110; Ex. 1015 (*Hoffman*), 1 ("virtually all real world classification systems have complex hierarchical structure"). The same holds true for *Deliu*'s other classification techniques. *Dec.*, ¶¶111. *Deliu*'s k-NN and

Decision Trees classifiers would similarly be improved by implementing a hierarchy or taxonomy among the labels. *Id.*, ¶¶111.

A POSITA would have had a reasonable expectation of success modifying *Deliu* to include hierarchical labels as taught by *Pendar*. *Id.*, ¶¶112-113. *Deliu* already pre-labels posts into one of several different classes/labels. *Deliu*, 74-76 (Fig. 12). It would have been well within the skill of a POSITA to pre-label posts hierarchically. *Dec.*, ¶¶112-113. Classifying ground truth data into a hierarchical form was a well-known strategy in machine learning. *Id.* Thus, a POSITA would have had success implementing that known strategy in *Deliu*'s methodology. *Id.*

4. Claim 1(a)(2): access data related to a discussion topic for classification from a deep web forum;

Deliu teaches accessing data from the deep web forum, Nulled.IO. *Supra*, Claim 1(a)(1). The Nulled.IO forum is “organized in **topics**, where each topic can have hundred[s] or thousands of posts.” *Deliu*, 104. “The main topics related to security in the Nulled.IO forum are...leaked credentials, web and network attacks (SQL injection, DDOS), spamming, and different malware types (Trojans, keyloggers, crypters, RATs, etc)[.]” *Id.*, 105; *see also id.*, 119-121 (Appendix A2 providing tables of “[t]opics of posts from Nulled.IO forum”). Using “Classification Methods[.]” *Deliu* classifies forum posts into categories. *Id.*, 74-76, Fig. 12.

5. ***Claim 1(a)(3): extract a set of features from the data by assigning word vectors or paragraph vectors to the data; and***

Deliu teaches extracting features from the Nulled.IO dataset using “Bag-of-Words” and “n-gram” feature extraction techniques. *Deliu*, 65-66, 85. These features are shown in Table 19 below:

Features	k-NN	Decision Trees	SVM
word (uni+bi)-grams	37.48	96.41	96.93
character trigrams	68.07	95.96	98.62
character(bi+tri)- grams	81.36	95.98	98.59
bag-of-words	66.76	96.45	97.27

Table 19: Classification accuracy of three classifiers on the multi-class dataset: k-Nearest Neighbors, Decision Trees, and Support Vector Machines

Id., 86 (Table 19, annotated); *see also id.*, 87 (Table 21).

Deliu teaches that both bag-of-words and n-gram assign word vectors. *Id.*, 65-66. Bag-of-words is “one of the most common methods for text classification,” outputting feature vectors representing “word frequency within each document[.]” *Id.*, 65. “That is, the dimensionality of the feature vectors is equal to the number of unique words in all documents in the dataset.” *Id.* (providing example word vectors for an illustrative dataset consisting of three documents). N-gram similarly assigns word vectors but considers “word order[.]” *Id.*, 66 (“word order is considered in a local context by n-gram features, which are defined as a sequence[] of *n* consecutive words/characters”); *Dec.*, ¶¶117-118.

6. ***Claim 1(a)(4): apply a machine classifier to the set of features to generate a prediction list of tags for classifying the discussion topic, wherein the prediction list includes a prediction probability value for each tag of the plurality of tags; and***

Deliu applies machine classifiers, such as SVM, k-NN, Decision Tree, and CNNs, to the extracted features to generate labels for each discussion post in the Nulled.IO forum. *Deliu*, 78-87 (“6.3 Classification of forums posts”); *see also id.*, 76 (Figure 12 illustrating distribution of labeled posts). *Deliu* “measure[s] the performance of traditional classifiers” in its datasets using different features. *Id.*, 80. This is summarized in Table 19 below:

Features	k-NN	Decision Trees	SVM
word (uni+bi)-grams	37.48	96.41	96.93
character trigrams	68.07	95.96	98.62
character(bi+tri)-grams	81.36	95.98	98.59
bag-of-words	66.76	96.45	97.27

Table 19: Classification accuracy of three classifiers on the multi-class dataset: k-Nearest Neighbors, Decision Trees, and Support Vector Machines

Id., 86 (Table 19), 87 (Table 21 comparing results of different “Classifier[s]” and “Features”); *see also id.*, 35 (“Most of the existing machine learning algorithms are able to build predictive models or infer intrinsic structural information **only when the data is presented in the form of feature vectors.**”), 41 (“Since machine learning algorithms are able to build predictive models only **from the features extracted** from raw data, their success is closely related to the quality of the

features”). *Deliu* therefore applies machine classifiers to the set of features to generate tags for classifying the discussion topic. *Dec.*, ¶122; Ex. 1016 (*Deliu* Paper), 3 (“The multinomial data set has a single class/label for the data from the *irrelevant* posts and then several different classes/labels for the *relevant* posts”).

Deliu does not expressly discuss generating “a prediction list of tags...wherein the prediction list includes a prediction probability value for each tag of the plurality of tags[.]” *Pendar* teaches this.

Pendar teaches a “tagging apparatus” that “assigns labels” by executing queries on documents with machine readable text. *Pendar*, 14:19-27, 15:27-62, Fig. 5. “[E]ach set of queries [is] associated with a label” such that multiple labels can be applied to a document. *Id.*, 14:25-28, 14:34-36. To apply “one or more labels[.]” *Pendar* calculates a prediction score for each query. *Id.*, 14:37-47. “One method of scoring may include summing each of the zero or more weights associated with positive queries.” *Id.*, 14:40-42. *Pendar* compares these scores to a “threshold” and “for each score meeting the threshold score, [*Pendar*] assign[s] one or more labels associated with the score to the document.” *Id.*, 14:58-67. “[I]f a document retrieved from an electronic source has summed scores exceeding a pre-defined threshold for ‘pizza’ and ‘spaghetti’, each label may be associated with the document[.]” *Id.*; see also *id.*, 15:27-16:3.

Pendar therefore predicts multiple labels for a document (i.e., “generate a prediction list of tags for classifying the discussion topic”). *Id.*, 14:28-36 (describing a first, second, and n^{th} label can be applied); *Dec.*, ¶¶125-126. This prediction list includes a prediction probability value for each tag of the plurality of tags. *Pendar*, 14:37-67 (a label is applied by summing the scores of each positive query such that when a document receives a summed score above a predetermined threshold, the label is assigned); see also *id.*, 9:56-62; *Dec.*, ¶¶125-126. The “summed” “scores” for each label are “values” that are compared to a threshold to predict a label and are therefore prediction probability values as claimed. *Dec.*, ¶¶125-126.

A POSITA would have been motivated to modify *Deliu*’s classifiers to implement *Pendar*’s multi-label querying techniques to facilitate assigning more than one label to each document. *Dec.*, ¶¶128-133. *Deliu* explains that categories selected for his study were “simple enough” for effective classification, but acknowledged that “more complex feature space topologies” exist. *Deliu*, 103 (discussing classifier performance “starts to get better than traditional approaches only when the dataset contains millions of samples”); Ex. 1016 (*Deliu* Paper), 7-8. Complex topologies would benefit from classifying a document into multiple categories to account for overlap where an item can belong to several categories at once, providing a more realistic classification for the dataset that accounts for ambiguity in the data. *Dec.*, ¶¶128-130. Additionally, *Pendar*’s predicted labels and

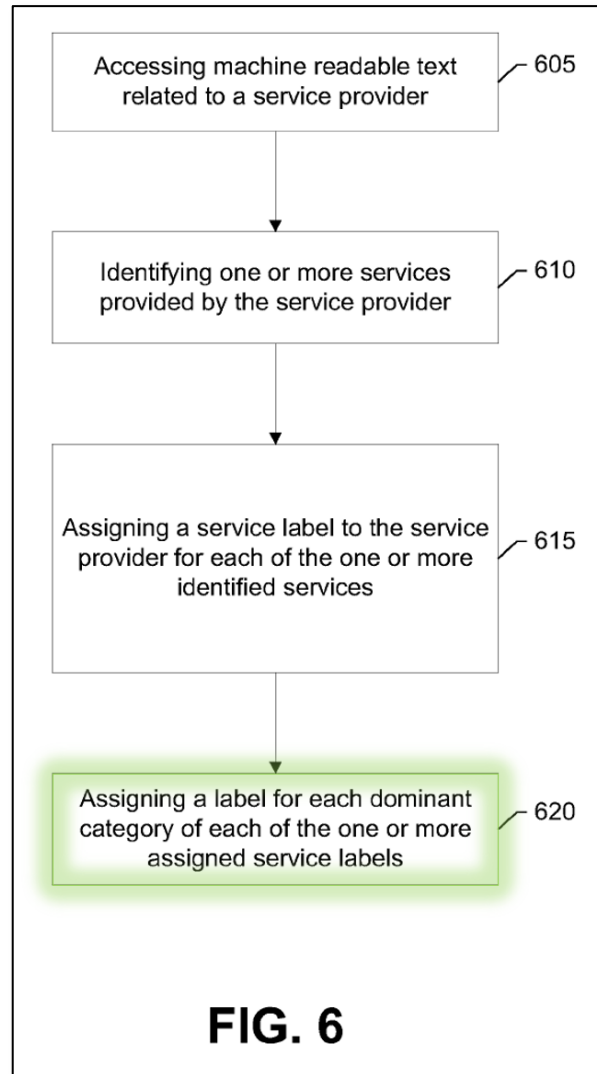
scores would have provided important information regarding the “informativeness” of each label. *Pendar*, 12:44-49, Fig. 3. A POSITA would have understood this data would have enhanced *Deliu*’s classifiers to more effectively differentiate and classify documents in more complicated feature space topologies, as contemplated by *Deliu*. *Dec.*, ¶¶131.

A POSITA would have had a reasonable expectation of success modifying *Deliu*’s classifiers to include *Pendar*’s similar classification techniques. *Dec.*, ¶¶132-133. In the modified system, *Deliu* classifiers would run *Pendar*’s “learned” queries for each category for a particular document, obtaining a score for each predicted label. *Id.* *Deliu* would then run its traditional classification techniques for each predicted label to assign the labels that exceed a threshold. *Id.* (discussing running multiple binary classifications for each predicted label to generate multiple labels for a document). *Deliu* would thus generate multiple predicted labels and scores for a document. *Id.* (noting multiple labels account for hierarchies present in the data); Ex. 1015 (*Hoffman*), 3 (“2.5 Hierarchical Multilabel Classification” describing coupling binary classifiers for multi-label settings to accomodate hierarchical classifications).

7. ***Claim 1(a)(5): add all parent tags associated with a tag of the plurality of tags to the prediction list based on a comparison between the prediction probability value for the tag and a first predetermined threshold value.***

Deliu-Pendar teaches generating multiple predicted tags (i.e., *prediction list*) and summed scores for those tags (i.e., *prediction probability value*) for posts in the Nulled.IO dataset. *Supra*, Claim 1(a)(4). *Pendar* further teaches an additional label assignment technique where “dominant categories” are assigned for each assigned label (i.e., *add all parent tags*). *Pendar*, 15:46-16:3. *Pendar* appreciates “one or more labels may be determined to be related to another label in the hierarchy.” *Id.*, 15:52-53. “For example, the label ‘burgers’ may be associated to ‘traditional American cuisine’ and ‘meat’.” *Id.*, 15:53-55. This is exemplary of *Pendar*’s “predefined hierarchical structure that includes at least 3 levels.” *Id.*, 15:55-57. *Pendar* teaches in “such cases, a label may be associated with one or more labels from a first category and the one or more labels from the first category are then associated to a label from a second category.” *Id.*, 15:57-62 (“the label ‘lobster’ may be associated to ‘seafood’. ‘Seafood’ may then be associated to ‘food and drink’”).

Continuing the example, “in an instance in which a label ‘lobster’ is determined, the label ‘lobster’ may be assigned, a dominant label ‘seafood’ may be assigned, and a dominant label ‘food and drink’ may be assigned.” *Id.*, 15:66-16:3. This process is shown below in Figure 6:



Id., Fig. 6 (annotated). In other words, *Pendar* teaches applying all dominant labels (i.e., *parent tags*) for tags that have been assigned a label. *Dec.*, ¶¶136-138. In the modified system, *Deliu-Pendar* assigns labels based on its classifiers comparing scores to thresholds. *See, e.g., Pendar*, 14:48-51 (“checking” the summed score “to a threshold score”), 14:51-57 (if the summed score “exceeds a predetermined threshold[,]” then the label is applied to the document). *Deliu-Pendar* therefore teaches adding all dominant labels associated with a label to the prediction list for

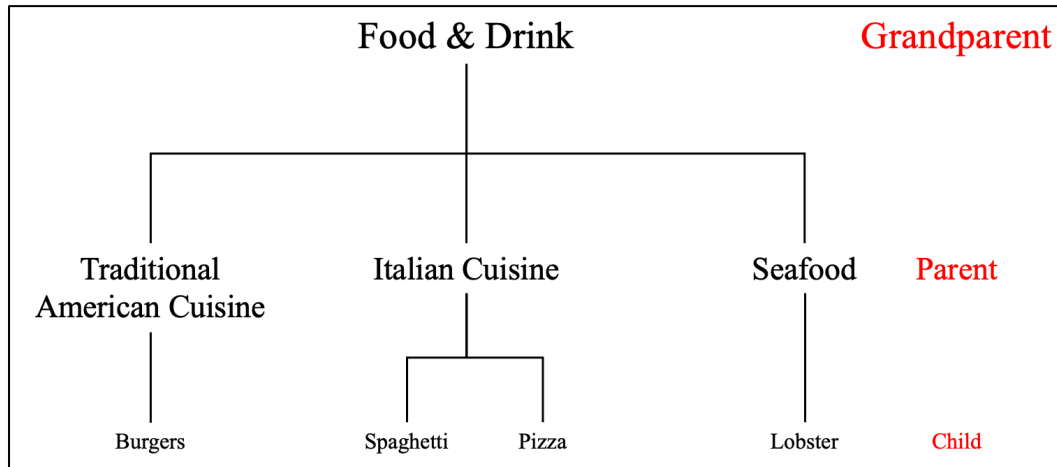
labels that exceeded a predetermined threshold (i.e., *add all parent tags associated with a tag of the plurality of tags to the prediction list based on a comparison between the prediction probability value for the tag and a first predetermined threshold value*). *Dec.*, ¶¶136-138.

For the same reasons above in Claim 1(a)(4), a POSITA would have been motivated to improve *Deliu's* machine classifier with *Pendar's* hierarchical classification techniques. *Dec.*, ¶¶140-144. *Pendar's* techniques include adding all labels for each “dominant category” (i.e., *parent tags*) of a classified label. *Pendar*, 15:46-16:3. This modification would help reduce false negative predictions of higher-level tags by predicting the most specific lower-level tag and auto-populating the higher-level tags. *Dec.*, ¶141. This increases classifier performance and reduces missed labels. *Id.* It also helps reduce semantically inconsistent predictions that contradict the natural hierarchy in the dataset. *Id.* Further, this would be applying a known technique (e.g., *Pendar's* method of assigning labels to each dominant category of that label) to a known device ready for improvement (e.g., *Deliu's* machine learning methods) to yield predictable results (e.g., improved classifier performance). *KSR Int'l Co. v. Teleflex Inc.*, 550 U.S. 398, 416 (2007); *Dec.*, ¶¶142. Therefore, a POSITA would have found it obvious to modify *Deliu's* classifier to add all dominant labels of a classified label for accurate and thorough data classification. *Dec.*, ¶¶142.

A POSITA would have had a reasonable expectation of success modifying *Deliu* to add all dominant labels of a classified label. *Dec.*, ¶¶143-144. This is a simple and known technique for improving classification that would have leveraged the natural hierarchy in *Deliu*'s datasets, utilizing the same hardware and requiring minor modifications to software that would have been well within the skill of a POSITA at the time. *Id.*, ¶¶143-144; Ex. 1008 (*Ramírez-Corona*), 1 (“In hierarchical classification, an example that belongs to certain class automatically belongs to all its superclasses”).

- B. *Claim 2 - The computer-implemented system of claim 1, wherein a predetermined tag hierarchy is divided into parent tags and child tags, wherein a topic represented by a child tag is representative of a subset of a topic represented by a parent tag, wherein each child tag corresponds to a particular parent tag, and each tag is assigned a corresponding probability value.***

Deliu-Pendar teaches the computer-implemented system of claim 1. *Supra*, Claim 1. *Deliu-Pendar* also teaches its dataset from the Nulled.IO hacker forum includes a predetermined tag hierarchy. *Id.*, Claim 1(a)(1). Namely, *Pendar* teaches a tag hierarchy with at least three levels. *Pendar*, 8:32-36, 15:1-23, 15:46-62. This tag hierarchy is illustrated in the exemplary annotation below:



Dec., ¶146-147. As shown, the hierarchy is divided into parent tags and child tags, where child tags are a subset of a topic represented by the parent tag. *Id.* For example, “the label ‘lobster’ [(*child tag*)] may be associated to ‘seafood’ [(*parent tag*)].” *Pendar*, 15:60-62.

In other examples, the parent tag, “Italian cuisine[.]” corresponds with the child tags “spaghetti” and “pizza[.]” *Pendar*, 15:6-15. A POSITA would have understood that “spaghetti” is *a topic represented by a child tag* and that it is *representative of a subset of a topic represented by a parent tag*—Italian cuisine. *Dec.*, ¶148. Each of the tags “spaghetti” and “pizza” correspond to a particular parent tag, Italian cuisine. *Dec.*, ¶148; *see also Pendar*, 15:53-56 (“the label ‘burgers’ may be associated to ‘traditional American cuisine’ and ‘meat’”), 15:55-62 (“[T]he label ‘lobster’ may be associated to ‘seafood’. ‘Seafood’ may then be associated to ‘food and drink’”).

Finally, for the same reasons discussed above in Claim 1(a)(4), *Pendar* teaches *each tag is assigned a corresponding probability value*. *Supra*, Claim 1(a)(4); *Pendar*, 14:34-36 (“an associated label of a first service, a second service and an n^{th} service may be applied”). Specifically, *Pendar* teaches labels are applied by summing scores of each query returning a positive result for a document such that when a document receives a summed score above a predetermined threshold, the labels are applied. *Pendar*, 9:63-10:7, 14:34-57. *Pendar* teaches the document is “assigned” the summed score. *See, e.g., id.*, 14:42-47 (“the document may be **assigned** a score of ‘0.25’ for the first [label] and ‘2.2’ for the second [label]”). In the case of multiple labels for a single post, values are assigned for each applied label. *Dec.*, ¶¶149-150. This includes labels following a natural hierarchy such that each of a child label and a parent label are applied to the post with assigned values. *Id.*, ¶¶149-150 (providing example of “Security” parent label and “Trojans” child label both applied to a post with assigned values above a threshold).

C. *Claim 4 – The computer-implemented system of claim 1, wherein the processor is configured to add all parent tags of the plurality of tags corresponding to a child tag if a probability value associated with the parent tag is above the first predetermined threshold value.*

As discussed in Section III.C.2 above, Petitioner proposes the Board correct a typographical error in Claim 4 and construe the claim consistent with the specification and file history to require adding all parent tags “if a probability value

associated with the ~~parent~~ [child] tag is above the first predetermined threshold[.]”

Supra, Sec. III.C.2.

Deliu-Pendar teaches the computer-implemented system of claim 1. *Supra*, Claim 1. Claim 4 specifies the probability value of the child tag must be “above” the threshold value. *Pendar* teaches this. *Pendar* teaches “checking” the summed score of a child label “to a threshold score” to determine whether that label should be assigned. *Supra*, Claim 1(a)(5); *Pendar*, 14:48-51. If the summed score “**exceeds** a predetermined threshold,” then the label is applied. *Pendar*, 14:51-57; *see also id.*, 10:6-7. *Pendar* further teaches that if a document is assigned a label, then the label for each “dominant category of **that label**” is also assigned. *Supra*, Claim 1(a)(5); *Pendar*, 15:63-16:3, Fig. 6. In other words, when a document is assigned a sub-dominant label (i.e., *child tag*) because its summed score exceeds a threshold, then the label(s) of the dominant category (i.e., *parent tags*) are also assigned. *Pendar*, 15:63-16:3, Fig. 6. *Deliu-Pendar* therefore teaches adding all dominant labels (parent tags) if a probability value associated with the sub-dominant label (child tag) is above the first predetermined threshold. *Dec.*, ¶¶153-155.

D. Claim 5 - The computer-implemented system of claim 1, wherein the set of features are extracted using textual feature extraction to identify tags associated with topic titles of the data.

Deliu-Pendar teaches the computer-implemented system of claim 1. *Supra*, Claim 1. *Deliu* teaches extracting features using the textual feature extraction

algorithms “Bag-of-Words” and “n-gram[.]” *Id.*, Claim 1(a)(3) (noting Bag-of-words is “one of the most common methods for text classification” for outputting word vectors and N-gram similarly assigns word vectors but considers “word order”); *Deliu*, 65-66 (describing Bag-of-Words as “[o]ne of the conventional methods for generating text features... features are some sort of word frequency within each document of a corpora... the feature vectors is equal to the number of unique words in all documents in the dataset” **and** describing n-gram features as “sequences of ***n* consecutive words/characters[.]**”).

As discussed above in Claim 1(a)(1), *Deliu* teaches hacker forum posts labeled as DDOS, Trojans, SQL Injections, etc. *Supra*, Claim 1(a)(1); *Deliu*, 74-76, Fig. 12. These labels correspond to topics in the forum. *Deliu*, 77 (Table 7). *Deliu* further teaches applying Latent Dirichlet Allocation (LDA) for “validation of the labeling process[.]” *Id.*, 91; *see also id.*, 31 (providing background on LDA “topic modeling”). Table 26 shows results from validating the SQL injection classification (i.e., tag, highlighted in **orange**). The LDA technique identifies topics (i.e., topic titles, highlighted in **green**) associated with the SQL injection tag. *Id.*, 91.

Table 26: Topics from SQL injection documents

Topic #1	Topic #2	Topic #3	Topic #4	Topic #5	Topic #6	Topic #7	Topic #8	Topic #9	Topic #10
sqli	man	ty	program	explanation	thanks	interested	sentry	php	sqli
sql	thanks	sql	sqli	website	sqli	love	mba	58	dumper
use	injectable	injection	dumper	vulnerability	sql	hits	vulnerabilities	http	hide
dorks	dorks	xss	does	based	injection	lol	idk	www	thanks
dumper	sql	possible	kind	explain	tutorial	hq	mysql	id	need
injection	sqli	error	league	crack	site	sqli	database	select	http
learn	hide	injections	net	help	work	cracking	sqli	91	com
know	working	use	source	understand	hack	gyazo	rdp	union	got
want	urls	site	lol	experienced	learning	dorks	proxies	999999	dorks
good	dumper	products	error	hacking	lot	dumping	errors	com	download

Id., Table 26 (annotated); *see also id.*, 119-121 (showing topic titles associated with the other described tags).

E. Claim 10 - The computer-implemented system of claim 1, wherein the set of features are extracted from the data using a neural network, wherein the neural network is configured to learn a continuous distributed vector representation for documents and words within the data.

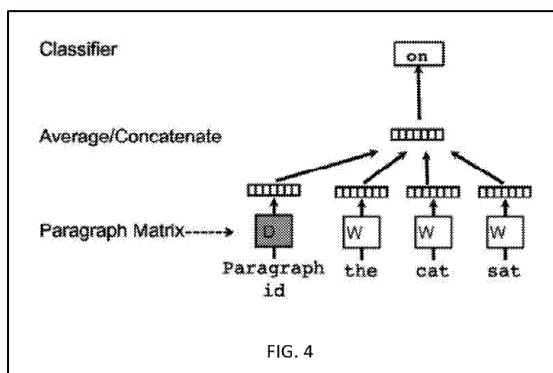
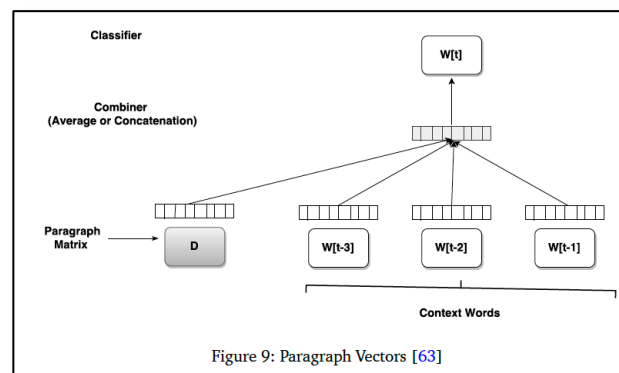
Deliu-Pendar teaches the computer-implemented system of claim 1. *Supra*, Claim 1.

Deliu teaches testing additional classification techniques on its Nulled.IO dataset, including using “Deep Neural Network” classifiers such as a “Convolutional Neural Network” (CNN). *Deliu*, 82. *Deliu*’s CNN classifier operates materially similar to its traditional approaches, and as modified by *Pendar*, would teach the limitations of claim 1 for the same reasons discussed with respect to *Pendar*’s other classifiers. *Dec.*, ¶162.

Deliu teaches several different vectorized input to its CNN, including “paragraph vectors[.]” *Deliu*, 65 (“The following is a brief overview of supervised

classification algorithms used[.]”), 66 (“Paragraph Vectors”), 82 (Table 13 “Paragraph Vectors”). Paragraph Vectors extract features using a neural network. *Id.*, 66 (discussing Paragraph Vectors follow the same principles as word embeddings). Word embeddings are a “**continuous vector representation model**” and consist “of a simple feed-forward **neural network**[.]” *Id.*, 36; *Dec.*, ¶163 (discussing the ’900 Patent’s similar description of PV-DM using a neural network). Paragraph Vectors parameterize both “the **entire text** (paragraph)” and “the vectors of **context words**” to predict the next word in the paragraph. *Deliu*, 66. Thus, the “paragraph vectors” used by *Deliu* are a continuous distributed vector representation, as claimed. *Dec.*, ¶164; Ex. 1019 (*Le*), 1 (Paragraph Vector is “an unsupervised framework that learns continuous distributed vector representations for pieces of texts”).

Indeed, *Deliu*’s paragraph vectors disclosure overlaps with the paragraph vector description in the ’900 Patent:



Dec., ¶¶165-166; '900 *Patent*, Fig. 4 (left); *Deliu*, Figure 9 (right). In the same way the “paragraph vectors” in the specification are extracted from the data using a neural network and represent continuous distributed vector representations for documents and words within the data, *Deliu*’s “paragraph vectors” teach the same. *Dec.*, ¶¶165-167.

F. Claim 12

1. Claim 12(Pre): A method, comprising:

To the extent the preamble is limiting, *Deliu* teaches a method to classify textual data from hacker forums. *Supra*, Claim 1(Pre); *Deliu*, 7 (“In this thesis, we explore the capabilities of Machine Learning methods in the task of locating relevant threat intelligence from hacker forums”).

2. Claim 12(a): configuring a processor for executing operations including:

Deliu teaches configuring a processor for executing operations with tools and software. *Supra*, Claim 1(a).

3. Claim 12(a)(1): accessing data associated with a deep web forum, the data defining a topic for classification;

Deliu teaches accessing data from the deep web Nulled.IO forum defining a discussion topic for classification. *Supra*, Claim 1(a)(2). The Nulled.IO forum is “organized in **topics**, where each topic can have hundred[s] or thousands of posts.” *Id.*; *Deliu*, 104. Using classification methods, *Deliu* classifies forum posts into categories. *Deliu*, 74-76, Fig. 12.

4. ***Claim 12(a)(2): extracting a set of features from the data as inputs for a machine classifier; and***

Deliu teaches extracting features from the Nulled.IO dataset using “Bag-of-Words” and “n-gram[.]” *Supra*, Claim 1(a)(3). “Bag-of-Words” and “n-gram” extract features to form word vectors that are used as inputs for *Deliu*’s k-NN, Decision Trees, SVM, and CNN classifiers. *Deliu*, 35 (“Most of the existing machine learning algorithms are able to build predictive models or infer intrinsic structural information **only when the data is presented in the form of feature vectors.**”), 41 (“Since machine learning algorithms are able to build predictive models only **from the features extracted** from raw data, their success is closely related to the quality of the features[.]”), 80 (Table 10), 86 (Table 19).

5. ***Claim 12(a)(3): apply a machine classifier to the set of features to generate a prediction list of tags for classifying the topic, wherein the prediction list includes a prediction probability value for each tag of the plurality of tags; and***

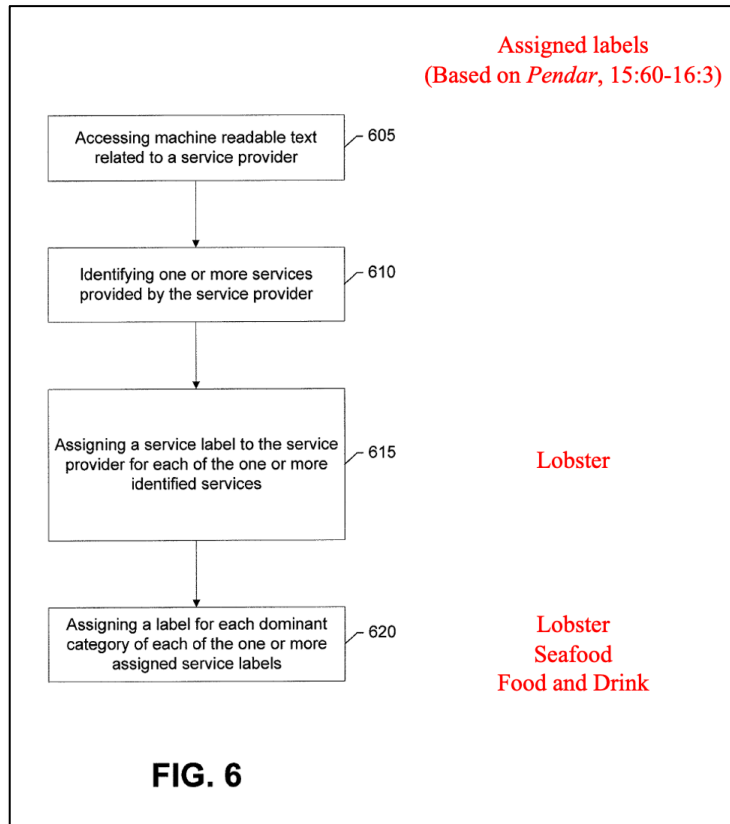
Deliu-Pendar teaches this limitation. *Supra*, Claim 1(a)(4) (materially identical claim language).

6. ***Claim 12(a)(4): adding all parent tags associated with a tag of the plurality of tags to the prediction list based on a comparison between the prediction probability value for the tag and a first predetermined threshold value.***

Deliu-Pendar teaches this limitation. *Supra*, Claim 1(a)(5) (materially identical claim language).

- G. *Claim 13 - The method of claim 12, further comprising: outputting an update to the prediction list of tags, a total number of the tags of the prediction list of tags being different from the update to the prediction list of tags to maintain a predetermined tag hierarchy.***

Deliu-Pendar teaches the computer-implemented system of claim 12. *Supra*, Claim 12. *Deliu-Pendar* teaches generating a list of labels associated with a document (*Pendar*, Fig. 6 (step 615) discussed above in Claim 1(a)(4)) and adding labels for each dominant category of the assigned labels (*Pendar*, Fig. 6 (step 620) discussed above in Claim 1(a)(5)). In *Deliu's* modified method, the process of adding dominant labels “updates” the prediction list of labels assigned to a document. *Dec.*, ¶¶178-180. Specifically, after Step 615 in Figure 6, a document has, for example, a single label of “lobster[.]” *Pendar*, 15:57-62. After Step 620, two additional labels of “seafood” and “food and drink” have been added. *Id.*, 15:66-16:3. Thus, there are more assigned labels (i.e., *a total number of the tags of the prediction list of tags being different from the update to the prediction list of tags*) after the dominant hierarchical labels are added. *Dec.*, ¶¶178-180. This process is illustrated in annotated Figure 6 below:



Dec., ¶¶178-180; *Pendar*, Fig. 6 (annotated), 15:60-16:3. Adding the labels “seafood” and “food and drink” follows *Pendar*’s service taxonomy discussed above in Claim 1(a)(1). (i.e., *to maintain a predetermined tag hierarchy*). *Dec.*, ¶180.

VI. GROUNDS 2 AND 3: CLAIM 3 IS OBVIOUS OVER *DELIU*, *PENDAR*, AND *FRANK* OR ALTERNATIVELY, OVER *DELIU*, *PENDAR*, AND *RAMÍREZ-CORONA*

A. *Claim 3 - The computer-implemented system of claim 1, wherein the processor is configured to remove all child tags of the plurality of tags corresponding to a parent tag from the prediction list of tags where a probability value associated with the child tag⁴ is below a second predetermined threshold value.*

1. *Ground 2 – Deliu, Pendar, Frank*

As discussed in Section III.C.1 above, Petitioner proposes the Board correct a typographical error in Claim 3 and construe the claim consistent with the specification and file history to require that the “**parent**” tag’s probability value is checked against the second threshold, not the “child” tag. *Supra*, Sec. III.C.1 (“wherein the processor is configured to remove all child tags...where a probability value associated with the ~~child~~ [**parent**] tag is below a second predetermined threshold value”). *Deliu-Pendar-Frank* teaches this claim.

⁴ Petitioner’s Ground 2 challenges Claim 3 and proposes correcting a typographical error to align the claim with the specification and file history. *Supra*, Sec. III.C.1. Ground 3 challenges Claim 3 uncorrected. As written, there is no antecedent basis for “the child tag” probability value that is compared to a threshold. Petitioner therefore maps the prior art to one of the child tags of the “all child tags[.]” *Infra*, Ground 3.

Deliu-Pendar teaches the computer-implemented system of claim 1. *Supra*, Claim 1. Specifically, *Pendar* teaches a summed score (i.e., *a probability value associated with the ~~child~~ [parent] tag*). *Supra*, Claim 1(a)(4).

Frank teaches a critical value pruning technique that analyzes hierarchies from the bottom-up. *Frank*, 76. “When a subtree is considered for pruning, the value of the splitting criterion **at the corresponding node is compared to a fixed threshold**, and **the tree is replaced by a leaf if the value is too small.**” *Id.* In other words, a splitting criteria value of a node is compared to a fixed threshold (i.e., *a second predetermined threshold*). If the splitting criterion is below the threshold (i.e., *below a second predetermined threshold*), then the subtree is removed (i.e., *remove all child tags*) in place of a single leaf node. *Id.*; *Dec.*, ¶189-190.

A POSITA would have been motivated to combine *Deliu-Pendar*’s machine classifier with hierarchical classification techniques with the subtree pruning techniques taught by *Frank* for several reasons. *Dec.*, ¶¶192-198. *Deliu-Pendar* already teaches summing scores (i.e., *a probability value associated with the [parent] tag*) of each positive query associated with a label. *Pendar*, 14:40-67. Leveraging this summed score to further identify and remove lower-level tags that contradict the natural hierarchy of the dataset would improve classifier performance by reducing false positive predictions of the lower-level tags. *Dec.*, ¶194-195. This would be applying a known technique (e.g., *Frank*’s method of pruning a subtree)

to a known device ready for improvement (e.g., *Deliu-Pendar*'s hierarchical classification technique) to yield predictable results (e.g., improved classifier performance). *KSR Int'l Co. v. Teleflex Inc.*, 550 U.S. 398, 416 (2007); *Dec.*, ¶¶195-196.

A POSITA would have had a reasonable expectation of success modifying *Deliu-Pendar* to use its summed score to compare against *Frank*'s fixed threshold in deciding whether to remove all subtree labels. *Dec.*, ¶¶197-198. This is a simple and known technique for improving classification that would have leveraged the natural hierarchy in *Deliu-Pendar*'s datasets, utilizing the same hardware and requiring minor modifications to software that would have been well within the skill of a POSITA at the time. *Dec.*, ¶198.

2. Ground 3 – *Deliu, Pendar, Ramírez-Corona*

To the extent the Board does not correct the typographical error in claim 3 and instead reads this claim to require thresholding a “child” tag's probability value against a second threshold, *Deliu-Pendar-Ramírez-Corona* teaches this claim.

Deliu-Pendar teaches the computer-implemented system of claim 1. *Supra*, Claim 1. *Ramírez-Corona* teaches a “[h]ierarchical [m]ulti-label [c]lassification” system that uses “pruning” methods when “the information available is not sufficient to estimate the class of an instance at the lower levels in the hierarchy[.]” *Ramírez-Corona*, 1, 6. In these instances, the path of the predicted hierarchy is

pruned/truncated. *Id.*, 6-7. *Ramírez-Corona* teaches a bottom-up traversal of the hierarchy, where “[t]he hierarchy is traversed starting from the leaf nodes[.]” *Id.*, 7. Traversing from the leaf node to its parents, “when [a] pruning condition is met in one node, the traversing is stopped and the descendants of the node are pruned.” *Id.* This “pruning condition” occurs when “the **probability of the most probable child**” of a parent tag is less than a threshold (i.e., *a probability value associated with the child tag is below a second predetermined threshold value*). *Id.* If the probability of the child is less than the threshold, the pruning condition is met, and “traversing is stopped and the node and **its descendants are pruned.**” *Id.* *Ramírez-Corona* therefore *remove[s] all child tags of the plurality of tags corresponding to a parent tag when the probability value associated with the child tag of a particular parent tag is below a second predetermined threshold value.* *Dec.*, ¶¶202-204.

A POSITA would have been motivated to modify *Deliu-Pendar*’s hierarchical multi-level classifiers to implement *Ramírez-Corona*’s pruning technique to maintain coherence in the multi-level tags. *Dec.*, ¶¶206-213. *Pendar* provides a multi-level service taxonomy, for example, an “Italian cuisine” dominant label and “spaghetti” and “pizza” sub-dominant labels. *Pendar*, 15:1-17. Hierarchical topologies benefit from probability thresholds associated with sub-dominant labels to ensure compatibility with the dominant level. *Dec.*, ¶¶207-208. For example, a POSITA would have understood that a dominant level label of “Italian cuisine” is

incompatible with a sub-dominant label of “sushi[.]” *Dec.*, ¶209. Thus, a POSITA would have looked to the techniques of *Ramírez-Corona* to ensure that only compatible sub-dominant levels are used. *Dec.*, ¶210. Indeed, the combination would have appreciated the express benefits discussed in *Ramírez-Corona*, including providing “superior” classification results with pruning when needed compared to returning complete hierarchical paths for all nodes. *Ramírez-Corona*, 13; *Dec.*, ¶211.

A POSITA would have had a reasonable expectation of success modifying *Deliu-Pendar*’s multi-level classifiers to implement *Ramírez-Corona*’s pruning technique. *Dec.*, ¶¶212-213. *Pendar* already sums scores associated with a probability to predict a label. *Pendar*, 9:56-62; *Dec.*, ¶¶209-210. Modifying *Pendar* to additionally examine the probability of sub-dominant labels would have been a simple modification to software, something that would have been well within the skill of a POSITA at the time. *Dec.*, ¶¶212-213.

VII. GROUND 4: CLAIMS 6 AND 14 ARE OBVIOUS OVER *DELIU, PENDAR, AND CHAWLA*

- A. *Claim 6 - The computer-implemented system of claim 1, wherein a set of synthetic data samples are created for addition to the ground truth dataset to address class imbalance by creating a set of synthetic sample points along a line defined between a first minority feature vector and a second feature vector and adding the set of synthetic sample points to the ground truth dataset.***

Claim 6 as-issued is identical to the originally filed claim 6 in prosecution.

'900 *File History*, 30. During prosecution, the Examiner rejected then-pending claim 6 based on *Chawla's* teachings:

Chawla et al. is cited to disclose wherein a set of synthetic data samples are created for addition to the ground truth dataset to address class imbalance by creating a set of synthetic sample points along a line defined between a first minority feature vector and a second feature vector and adding the set of synthetic sample points to the ground truth dataset ("We generate synthetic examples in a less application-specific manner by operating "feature space" rather than "data space". The minority class is over-sampled by taking each minority class sample and introducing synthetic examples along the line segments joining any/ all of the k minority class nearest neighbors," Chawla et al., sec. 4.2, 1st para. See also Chawla et al., fig. 3.).

Id., 228-229. Applicant did not amend claim 6 or argue against *Chawla's* disclosures during prosecution. *Id.*, 278-294. Applicant amended the independent claims and (incorrectly) argued claim 6 depends from a valid independent claim. *Id.*, 293. Claim 6 then issued unamended. *Id.*, 295-302.

As discussed above, independent claim 1 should be found unpatentable in view of *Deliu-Pendar. Supra*, Claim 1. *Chawla* teaches the limitations of Claim 6.

Notably, *Chawla* teaches a near verbatim process for generating synthetic data samples as described in the '900 Patent's specification. The table below compares those disclosures:

<u>Chawla</u>	<u>'900 Patent</u>
<p>"4.2 SMOTE...The minority class is over-sampled by taking each minority class sample and introducing synthetic examples along the line segments joining any/all of the k minority class nearest neighbors. Depending upon the amount of over-sampling required, neighbors from the k nearest neighbors are randomly chosen. Our implementation currently uses five nearest neighbors...Synthetic samples are generated in the following way: Take the difference between the feature vector (sample) under consideration and its nearest neighbor. Multiply this difference by a random number between 0 and 1, and add it to the feature vector under consideration. This causes the selection of a random point along the line segment between two specific features."</p>	<p>"Synthetic Minority Over-Sampling Technique (SMOTE)...In SMOTE, the minority class is over-sampled by taking each minority class sample and introducing synthetic examples along the line segments joining all of the k minority class nearest neighbors. Depending upon the amount of over-sampling required, neighbors from the k nearest neighbors are randomly chosen. The present system currently uses five nearest neighbors. Synthetic samples are generated in the following way:</p> <ol style="list-style-type: none"> 1) Take the difference between the feature vector under consideration and its nearest neighbor. 2) Multiply this difference by a random number between zero and one. 3) Add it to the feature vector under consideration. <p>In this manner, a random point is synthetically introduced along the line segment between two specific features."</p>

Compare, Chawla, 8 with '900 Patent, 7:11-34.

A POSITA would have been motivated and would have found it obvious to implement *Chawla*'s SMOTE technique for dealing with class imbalance in *Deliu-Pendar*'s data classification system. *Dec.*, ¶¶222-228. *Deliu*'s machine learning

model for forum classification relies on labeled samples in a dataset. *Deliu*, 62. However, the number of samples can be limited, and *Deliu* describes a process for “[s]nowball sampling” to “increase the number of samples (like a snowball) from the existing samples” for situations where “‘recruiting’ of samples is difficult[,]” like identifying additional hacker forums. *Id.*, 62. While this increases the overall number of samples, “snowball sampling” does not address class imbalance. *Dec.*, ¶¶224-225. *Chawla* teaches a method of increasing the number of samples in an imbalanced dataset by synthetically generating samples to add to the dataset. *Chawla*, 1. A POSITA would have found it obvious to synthetically generate samples to add to a dataset when, as described by *Deliu*, adding non-synthetic samples is challenging for at least the benefits discussed in *Chawla*. *Dec.*, ¶226; *Chawla*, 32 (“the SMOTE approach can improve the accuracy of classifiers for a minority class...SMOTE provides more related minority class samples to learn from, thus allowing a learner to carve broader decision regions, leading to more coverage of the minority class”). Thus, synthetically generating samples would have overcome a known limitation to classifying hacker forum posts. *Dec.*, ¶226.

A POSITA would have had a reasonable expectation of success in modifying *Deliu-Pendar* with *Chawla*. *Deliu* already teaches “snowball sampling[,]” “a technique used to increase the number of samples (like a snowball) from the existing samples and it is used in situations where ‘recruiting’ of samples is difficult.” *Deliu*,

62. Implementing SMOTE as taught by *Chawla* would have only required software modifications to *Deliu*'s current over-sampling techniques, something well within the skill of a POSITA. *Dec.*, ¶¶227-228. Indeed, *Chawla* offers the pseudo-code for SMOTE. *Chawla*, 8-9.

B. *Claim 14 - The method of claim 12, further comprising generating a predetermined tag hierarchy based on a ground truth dataset, and leveraging semi-supervised learning to address class imbalance and increase accuracy of the machine classifier.*

Deliu-Pendar teaches generating a predetermined tag hierarchy based on a ground truth dataset. *Supra*, Ground 1, Claim 1(a)(1) (describing hierarchically labeling the ground truth dataset from the Nulled.IO hacker forum). The '900 Patent describes SMOTE as a semi-supervised learning technique. '900 Patent, 12:12-13 ("SMOTE semi-supervised learning"); *see also id.*, 11:51-12:13. *Chawla* describes the same SMOTE technique as described in the '900 Patent and, therefore, "leverage[es] semi-supervised learning[.]" *Supra*, Claim 6; *Dec.*, ¶220.

Chawla's SMOTE technique "address[es] class imbalance and increase[s] accuracy of the machine classifier." *Chawla*, Abstract ("An approach to the construction of classifiers from **imbalanced datasets** is described....our method...can achieve **better classifier performance**"), 32 ("The results show that the **SMOTE approach can improve the accuracy** of classifiers for a minority class."), *id.* ("[SMOTE] **improves performance**"), *id.* ("SMOTE provides more

related minority class samples to learn from...**leading to more coverage of the minority class.**").

For the same reasons discussed above in Claim 6, a POSITA would have been motivated and would have found it obvious to implement *Chawla*'s SMOTE technique for dealing with class imbalance in *Deliu-Pendar*'s classification system. *Dec.*, ¶¶222-228; *supra*, Claim 6. Furthermore, for the same reasons discussed *supra* Claim 6, a POSITA would have had a reasonable expectation of success making this modification. *Id.*

VIII. GROUND 5: CLAIM 7 IS OBVIOUS OVER *DELIU*, *PENDAR*, AND *XIAO*

- A. *Claim 7 - The computer-implemented system of claim 1, wherein the ground truth dataset is supplemented by, for every sample of the ground truth dataset in a minority class, adding top similar documents from the deep web forum information, as identified using an elastic search similarity score, and adding them to the ground truth dataset.***

Deliu-Pendar teaches the computer-implemented system of claim 1. *Supra*, Claim 1.

Deliu teaches “snowball sampling[,]” “a technique used to increase the number of samples [i.e., *adding [samples] to the ground truth dataset*] (like a snowball) from the existing samples and it is used in situations where recruiting of samples is difficult.” *Deliu*, 62 (internal quotations omitted). However, *Deliu* does

not discuss using “an elastic search similarity score” to add “top similar documents” to ground truth. *Xiao* teaches this.

Notably, *Xiao* teaches using the “BM25” elastic search technique to identify similar documents—the same elastic search technique the applicant described in the specification of the ’900 Patent. *Compare, Xiao*, 3-4 (“the system calculates the **BM25-based similarity** between the query q and each document”) *with* ’900 Patent, 7:54-8:15 (“The elastic search uses BM25 as the default ranking function to provide the relevance score to a given search query”). *Xiao* teaches that “BM25 is a widely adopted weighting method in information retrieval[.]” *Xiao*, 3. “In the BM25 weighting scheme, the input document is viewed as a query q containing a [series] of terms $\{t_1, t_2, \dots, t_m\}$. The System collects all the documents in which at least one term occurs. “Then the system calculates the BM25-based similarity between the query q and each document d in training data set[.]” *Id.*, 3-4. This calculation is an “elastic search similarity score[.]” *Dec.*, ¶¶237-238 (noting *Xiao*’s similarity equation is the same equation described in the ’900 Patent’s specification)⁵.

⁵ In fact, the Applicant appeared to copy and paste from a Wikipedia webpage for the specification’s discussion of the BM25 technique and surrounding equation. Ex. 1022 (https://en.wikipedia.org/wiki/Okapi_BM25).

A POSITA would have been motivated to modify *Deliu-Pendar* to implement *Xiao*'s BM25 elastic search to identify additional similar documents to add to the existing dataset. *Dec.*, ¶¶240-243. Dark web forums are known to include a large number of irrelevant posts, a comparatively small number of security-relevant posts, and an even smaller number of posts relevant to specific security categories. *Deliu*, 76, 93-94, Figures 12 and 15; *Dec.*, ¶240. A POSITA would have been motivated to modify *Deliu-Pendar* with *Xiao*'s BM25 elastic search to increase the size of the less-frequently discussed security categories to improve the classifier's performance in identifying and correctly classifying these documents. *Dec.*, ¶241; *Xiao*, 6 ("BM25" "outperform[s] other similarity calculation methods"). This would be applying a known technique (e.g., *Xiao*'s BM25 method of identifying similar documents) to a known device ready for improvement (e.g., *Deliu*'s classification system) to yield predictable results (e.g., adding top similar documents to a dataset). *KSR Int'l Co. v. Teleflex Inc.*, 550 U.S. 398, 416 (2007); *Dec.*, ¶241.

A POSITA would have had a reasonable expectation of success in modifying *Deliu-Pendar* to implement *Xiao*'s BM25 elastic search to identify additional similar documents to add to the existing dataset. *Dec.*, ¶¶242-243. BM25 is a well-known technique for identifying similar documents. *Xiao*, 3 ("BM25 is a widely adopted weighting method in information retrieval"). BM25 would function in conjunction with *Deliu*'s snowball sampling to increase the number of top ranked similar

documents in a dataset, utilizing the same hardware and requiring minor modifications to software that would have been well within the skill of a POSITA at the time. *Dec.*, ¶243 (discussing BM25 Wikipedia article).

IX. GROUND 6: CLAIMS 8 AND 9 ARE OBVIOUS OVER *DELIU*, *PENDAR*, AND *UKRAINCZYK*

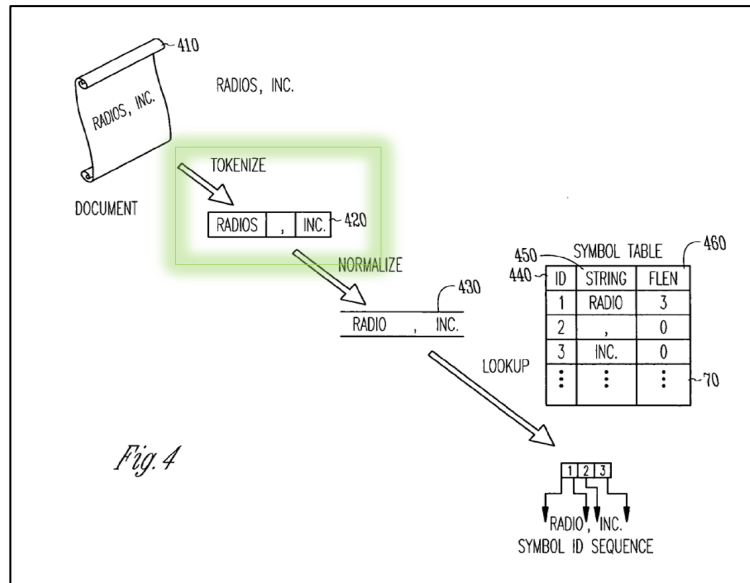
- A. *Claim 8 - The computer-implemented system of claim 1, wherein the data is parsed and segmented into tokens, and predetermined characters and predetermined common words are removed from the data.***

Deliu-Pendar teaches the computer-implemented system of claim 1. *Supra*, Claim 1.

Deliu teaches two data preprocessing sub-phases: preparation and cleaning. *Deliu*, 64. Regarding “preparation,” “[t]he objective of data preparation is the extraction of only relevant fields from the raw data and their storage in a[n] adequate format...This is especially needed when data is collected using web crawlers which store the data in a raw HTML format, and special text parsers are used to extract fields such as the content of the posts, title, author, date, etc.” *Id.* “On the other hand, the aim of data cleaning is to remove parts of the data which do not contribute to the goal (e.g. act as noise), and have a negative effect in the performance of given algorithms.” *Id.* (describing “Data cleaning methods” including “**Document Tokenization**”).

Deliu states that parsing data is “needed[,]” but *Deliu*’s data was already parsed. *Id.*, 64. As explained below, a POSITA would have been motivated to implement *Deliu*’s experiments on online hacker forums with data not already stored in relational databases like Nulled.IO. *Dec.*, ¶¶72-73, 252-256. For those online forums, *Deliu* would need to parse forum data into tokens and would look to known parsing techniques. *Ukrainczyk* describes one example of known methods for parsing a document into a plurality of unique tokens.

Ukrainczyk teaches classifying text into categories, where “tokens or features are...associated with each category.” *Ukrainczyk*, Abstract (“a document is parsed into a plurality of unique tokens with associated counts”). In classifying documents, a feature list is constructed including “a list of features automatically extracted from the training set without human intervention.” *Id.*, 5:62-63. Features from the feature list are classified into a symbol table after “the text is tokenized[.]” *Id.*, 7:54-55. *Ukrainczyk* describes a token as “a sequence of characters that makes up a single unit for processing.” *Id.*, 7:55-57. Tokens are full words (e.g., “Yahoo”), punctuation (e.g., “!”), and abbreviations (e.g., “Dr.”). *Id.*, 7:57-63. This tokenization process is illustrated below in Figure 4:



Id., Fig. 4 (annotated).

Ukrainczyk further teaches removing certain words and punctuation from consideration. For example, *Ukrainczyk* identifies stop words, including “the, in, of, at, to, etc.” *Id.*, 8:39-41. Stop words are assigned the attribute “Stop[,]” where “if set, the feature is a stop word—it is not counted during topic spotting. This bit is ignored during training[.]” *Id.*, 11:10-11. *Ukrainczyk* further teaches that “[s]top words, punctuation signs, XML tags, etc. are not passed to feature vector 100[,]” where the feature vector “represents the relevant topic-spotting features within a segment of text.” *Id.*, 12:14-15, 11:43-45.

A POSITA would have been motivated and would have found it obvious to implement *Ukrainczyk*’s methods for parsing and tokenizing documents and cleaning documents by removing predetermined characters and common words, such as punctuation and stop words, to facilitate *Deliu-Pendar*’s preprocessing sub-

phases on online hacker forum datasets. *Dec.*, ¶¶252-256 (describing *Deliu*'s process on live datasets). Tokenization simplifies text by breaking it into manageable tokens that aid in processing and analysis and reduce inconsistencies in language variations. *Id.*, ¶¶72-73, 254. Removing predetermined characters and predetermined common words reduces dataset size, speeds up processing, and reduces computation complexity. *Id.*, ¶¶72-73, 254. Removing predetermined characters and predetermined common words also reduces “noise[,]” which enhances classifier accuracy and improves feature extraction. *Id.*, ¶¶72-73, 255.

A POSITA would have had a reasonable expectation of success modifying *Deliu-Pendar* to parse and tokenize online hacker forum discussions and clean that tokenized data by removing predetermined characters and common words as taught by *Ukrainczyk* at least because *Deliu* already teaches both parsing/tokenizing and data cleaning techniques. *Dec.*, ¶256 (describing Table 5 in *Deliu*). This modification, therefore, would have required simple programming changes to the techniques already present in *Deliu*. *Id.*, ¶256.

B. *Claim 9 - The computer-implemented system of claim 1, wherein words within the data are reduced to their root form using stemming and lemmatization procedures.*

Deliu-Pendar teaches the computer-implemented system of claim 1. *Supra*, Claim 1.

As discussed above in Claim 8, *Deliu* teaches preprocessing hacker forum data. *Supra*, Claim 8. While *Deliu* does not expressly discuss stemming and lemmatization procedures during preprocessing, *Ukrainczyk* teaches this.

Ukrainczyk teaches a stemmer used to “return[] the **stem or lemma** of the word.” *Ukrainczyk*, 8:7-9. When finding word matches, *Ukrainczyk* analyzes “morphological differences” between two words. *Id.*, 7:35-40, 12:10-12 (“The same two-comparison mechanism is used to check for morphological matches when a feature’s Stemming attribute is set.”). Indeed, *Ukrainczyk* considers both “plural/singular [differences] and verb inflection[.]” *Id.*, 7:35-40. *Ukrainczyk* further “relates synonymous terms.” *Id.*, 7:43-44.

A POSITA would have understood *Ukrainczyk* teaches or suggests reducing words to their root form using stemming and lemmatization procedures consistent with the disclosures of the ’900 Patent. *Dec.*, ¶¶74-76, 259-262. The ’900 Patent describes “[l]emmatization” as “the process of grouping together different inflected forms of a word so they can be analyzed as a single item.” ’900 Patent, 9:7-9. Regarding “stemming,” the ’900 Patent states the “stemming process reduces the derived words—“caresses”, “ponies” and “cats” to “caress”, “poni”, and “cat” respectively.” ’900 Patent, 9:12-14. *Ukrainczyk* similarly describes reducing words in a dataset to their “normal form.” *Ukrainczyk*, 8:4-6 (“The normalization step takes a token or tag and returns the normal form of the token/tag.”). *Ukrainczyk* gives

similar examples: “carrying” is reduced to “carry,” and “dogs” to “dog.” *Ukrainczyk*, 8:9-11. *Dec.*, ¶¶259-262. In this way, *Ukrainczyk* “return[s] the **stem or lemma** of the word” using stemming and lemmatization procedures as claimed. *Ukrainczyk*, 8:7-9.

A POSITA would have been motivated and would have found it obvious to implement *Ukrainczyk*’s stemming and lemmatization data preprocessing techniques in *Deliu-Pendar*’s system to improve classification. *Dec.*, ¶¶264-266. Stemming and lemmatization reduce words to their “root forms,” which standardizes vocabulary, reduces data dimensionality which improves classifier efficiency, and improves information retrieval by grouping morphologically related words. *Dec.*, ¶265. These techniques help handle words with complex inflections or irregular forms, including those that would commonly be found in hacker websites. *Dec.*, ¶265.

A POSITA would have had a reasonable expectation of success modifying *Deliu-Pendar*’s preprocessing to further implement stemming and lemmatization because *Deliu* already teaches preprocessing its datasets, and stemming and lemmatization are widely known and routinely used preprocessing techniques with commonly known benefits. *Dec.*, ¶266.

X. GROUND 7: CLAIM 11 IS OBVIOUS OVER *DELIU*, *PENDAR*, AND *NUNES*

- A. *Claim 11 – The computer-implemented system of claim 1, further comprising: a crawler in operable communication with the processor and operable to traverse a plurality of deep web websites to retrieve Internet documents forming at least a portion of the deep web forum information.***

Deliu-Pendar teaches the computer-implemented system of claim 1. *Supra*, Claim 1.

Deliu teaches a processor. *Supra*, Claim 1(a). *Deliu* further teaches that data collection from hacker forums “demands” web crawlers. While *Deliu* does not expressly teach “a crawler in operable communication with the processor” because *Deliu* uses leaked forum data that already exists in a relational database, *Deliu* suggests web crawlers for use in live data collection from hacker forums. *Id.*, 62; *Dec.*, ¶270. *Deliu* does not provide details regarding specific deep/darknet web crawlers for use by its classification system.

*Nunes*⁶ teaches a crawler for darknet and deepnet mining of cyber threat intelligence. *Nunes*, Title (“**Darknet and Deepnet Mining for Proactive**

⁶ *Nunes* is a publication with two authors that are also inventors listed on the face of the '900 Patent. *Nunes* expressly describes a crawler that operates nearly verbatim to the claim language. Therefore, the inventors of the '900 Patent dedicated the concepts claimed in claim 11 to the public back in 2016 when *Nunes* published.

Cybersecurity Threat Intelligence”), Abstract (“In this paper, we present an operational system for cyber threat intelligence **gathering from various social platforms on the Internet** particularly sites on the **darknet and deepnet**.”). *Nunes* uses a crawler “to **traverse the website** and **retrieve HTML documents**.” *Id.*, 2; *see also id.*, Fig. 1. *Nunes* crawls a number of websites, including ten darknet/deepnet marketplaces and two darknet forums. *Id.*, 4 (“We consider 10 marketplaces to train and test our learning model.”), *id.* (“In addition to the **darknet/deepnet marketplaces** that we have already discussed, there are also numerous **darknet forums** on which users discuss malicious hacking related topics.”), *id.* (“We performed evaluation on **two such English forums**.”).

A POSITA would have been motivated to implement *Nunes*’ deepnet crawler in *Deliu-Pendar* to obtain live hacker online forum data for classification. *Dec.*, ¶¶274-281. While *Deliu* did not use a crawler because the leaked Nulled.IO dataset was already publicly available, *Deliu* acknowledges that the research “has a commercial development potential.” *Deliu*, 107. For example, *Deliu* envisions extracting cyber-threat intelligence from other hacker forums that have not been leaked to the Clearnet. *Id.*, 108 (“We are confident that relevant cyber-threat intelligence can also be extracted from hacker forums that use other languages, such as Russian and Chinese...So the scalability of our methods should be tested by extending them to other forums and platforms”); *Dec.*, ¶¶275-276 (discussing other

forums would not be found in leaked datasets). A POSITA would have understood that extending the machine learning classification techniques of *Deliu* to other non-leaked forums would “demand” a crawler to obtain the requisite forum data. *Deliu*, 7; *Dec.*, ¶276. A POSITA would have understood the crawler would have reduced human resources required to identify new forum information, something especially important for *Deliu*’s intention to scale techniques to a commercial platform. *Dec.*, ¶277.

A POSITA would have had a reasonable expectation of success implementing *Nunes*’s deepnet crawler to crawl online deepnet hacker forums to retrieve HTML documents. *Deliu* already acknowledges the need for data collection of “raw data” in constructing a dataset. *Deliu*, 61 (Figure 8, “Raw Data”), 62 (“The first phase of analyzing open sources for threat intelligence is the identification and collection of the data.”). And while *Deliu* has access to a leaked dataset, a POSITA would have expected success in obtaining *Deliu*’s dataset with a deepnet crawler as taught by *Nunes*. *Dec.*, ¶¶278-279. Web crawlers were prevalent at the time, and special-purpose deepnet crawlers were also well-known. *Dec.*, ¶¶77-79, 280-281; *Nunes*, 2 (discussing previous topic based crawlers and focused crawling). Implementing a known deepnet crawler in *Deliu-Pendar* would have required no more than the ordinary skill of a skilled artisan. *Dec.*, ¶281.

XI. CONCLUSION

For the foregoing reasons, Petitioner respectfully requests *inter partes* review of the Challenged Claims.

Respectfully submitted,

By: /s/ Paul R. Hart
Paul R. Hart, Reg. No. 59,646

COUNSEL FOR PETITIONER

XII. MANDATORY NOTICES UNDER 37 C.F.R. § 42.8(A)(1)

A. *Real Parties-In-Interest*

Petitioner and CrowdStrike Holdings, Inc. are the real parties-in-interest. 37 C.F.R. § 42.8(b)(1).

B. *Related Matters*

The '900 Patent is presently the subject of a patent infringement lawsuit filed against Petitioner in the United States District Court for the Western District of Texas, *Skysong Innovations, LLC v. CrowdStrike, Inc. and CrowdStrike Holdings, Inc.*, Case No. 7:25-cv-00040-DC-DTG. The '900 Patent has also been asserted against Fortinet, Inc. in the United States District Court for the Eastern District of Texas, *Skysong Innovations, LLC v. Fortinet, Inc.*, Case No. 2:25-cv-00098.

C. *Lead and Back-Up Counsel*

Petitioner provides the following designation and service information for lead and back-up counsel. 37 C.F.R. § 42.8(b)(3) and (b)(4).

Lead Counsel	Back-Up Counsel
Paul R. Hart (Reg. No. 59,646) Paul.Hart@eriseip.com PTAB@eriseip.com Postal and Hand-Delivery Address: ERISE IP, P.A. 717 17th St., Suite 1400 Denver, Colorado 80202 Telephone: (913) 777-5600 Fax: (913) 777-5601	Adam P. Seitz (Reg. No. 52,206) Adam.Seitz @eriseip.com PTAB@eriseip.com Postal and Hand-Delivery Address: ERISE IP, P.A. 7015 College Blvd., Suite 700 Overland Park, Kansas 66211 Telephone: (913) 777-5600 Fax: (913) 777-5601
	Hunter Horton, pro hac vice

	<p>Hunter.Horton@eriseip.com</p> <p>Postal and Hand-Delivery Address: ERISE IP, P.A. 7015 College Blvd., Suite 700 Overland Park, Kansas 66211 Telephone: (913) 777-5600 Fax: (913) 777-5601</p>
--	--

D. 37 C.F.R. § 42.8(b)(4) – Service Information

Please address all correspondence to the lead and back-up counsel as shown above. Petitioner consents to electronic service by e-mail at the e-mail addresses provided above.

CLAIM LISTING APPENDIX
U.S. Patent No. 11,275,900 for Claims 1-14

Claim Designation	Claim Language
Claim 1(Pre)	A computer-implemented system for improving classification of criminal activities arising from cyber environments, comprising:
Claim 1(a)	a processor configured to:
Claim 1(a)(1)	access a ground truth dataset generated from deep web forum information, the ground truth dataset defining a predetermined tag hierarchy including predetermined tag labels corresponding topics of the deep web forum information;
Claim 1(a)(2)	access data related to a discussion topic for classification from a deep web forum;
Claim 1(a)(3)	extract a set of features from the data by assigning word vectors or paragraph vectors to the data; and
Claim 1(a)(4)	apply a machine classifier to the set of features to generate a prediction list of tags for classifying the discussion topic, wherein the prediction list includes a prediction probability value for each tag of the plurality of tags; and
Claim 1(a)(5)	add all parent tags associated with a tag of the plurality of tags to the prediction list based on a comparison between the prediction probability value for the tag and a first predetermined threshold value.
Claim 2	The computer-implemented system of claim 1, wherein a predetermined tag hierarchy is divided into parent tags and child tags, wherein a topic represented by a child tag is representative of a subset of a topic represented by a parent tag, wherein each child tag corresponds to a particular parent tag, and each tag is assigned a corresponding probability value.

Claim Designation	Claim Language
Claim 3	The computer-implemented system of claim 1, wherein the processor is configured to remove all child tags of the plurality of tags corresponding to a parent tag from the prediction list of tags where a probability value associated with the child tag is below a second predetermined threshold value.
Claim 4	The computer-implemented system of claim 1, wherein the processor is configured to add all parent tags of the plurality of tags corresponding to a child tag if a probability value associated with the parent tag is above the first predetermined threshold value.
Claim 5	The computer-implemented system of claim 1, wherein the set of features are extracted using textual feature extraction to identify tags associated with topic titles of the data.
Claim 6	The computer-implemented system of claim 1, wherein a set of synthetic data samples are created for addition to the ground truth dataset to address class imbalance by creating a set of synthetic sample points along a line defined between a first minority feature vector and a second feature vector and adding the set of synthetic sample points to the ground truth dataset.
Claim 7	The computer-implemented system of claim 1, wherein the ground truth dataset is supplemented by, for every sample of the ground truth dataset in a minority class, adding top similar documents from the deep web forum information, as identified using an elastic search similarity score, and adding them to the ground truth dataset.
Claim 8	The computer-implemented system of claim 1, wherein the data is parsed and segmented into tokens, and predetermined characters and predetermined common words are removed from the data.
Claim 9	The computer-implemented system of claim 1, wherein words within the data are reduced to their root form using stemming and lemmatization procedures.

Claim Designation	Claim Language
Claim 10	The computer-implemented system of claim 1, wherein the set of features are extracted from the data using a neural network, wherein the neural network is configured to learn a continuous distributed vector representation for documents and words within the data.
Claim 11	The computer-implemented system of claim 1, further comprising: a crawler in operable communication with the processor and operable to traverse a plurality of deep web websites to retrieve Internet documents forming at least a portion of the deep web forum information.
Claim 12(Pre)	A method, comprising:
Claim 12(a)	configuring a processor for executing operations including:
Claim 12(a)(1)	accessing data associated with a deep web forum, the data defining a topic for classification;
Claim 12(a)(2)	extracting a set of features from the data as inputs for a machine classifier; and
Claim 12(a)(3)	apply a machine classifier to the set of features to generate a prediction list of tags for classifying the topic, wherein the prediction list includes a prediction probability value for each tag of the plurality of tags; and
Claim 12(a)(4)	adding all parent tags associated with a tag of the plurality of tags to the prediction list based on a comparison between the prediction probability value for the tag and a first predetermined threshold value.
Claim 13	The method of claim 12, further comprising: outputting an update to the prediction list of tags, a total number of the tags of the prediction list of tags being different from the update to the prediction list of tags to maintain a predetermined tag hierarchy.
Claim 14	The method of claim 12, further comprising generating a predetermined tag hierarchy based on a ground truth dataset, and leveraging semi-supervised learning to address class imbalance and increase accuracy of the machine classifier.

APPENDIX OF EXHIBITS

Exhibit 1001	U.S. Patent No. 11,275,900 to Patil et al. (“’900 Patent”)
Exhibit 1002	File History of U.S. Patent No. 11,275,900 (“’900 Patent File History”)
Exhibit 1003	Declaration of Erez Zadok, Ph.D. (“Dec.”)
Exhibit 1004	Isuf Deliu, <i>Extracting Cyber Threat Intelligence From Hacker Forums</i> , NTNU Norwegian University of Science and Technology, Doctoral Thesis Submitted June 2017, Carl Leichter, and Katrin Franke (“Deliu”)
Exhibit 1005	U.S. Patent No. 9,330,167 to Pendar (“Pendar”)
Exhibit 1006	Declaration of June Munford (“Munford Dec.”)
Exhibit 1007	Eibe Frank, <i>Pruning Decision Trees and Lists</i> , University of Waikato, Submitted January 2000 (“Frank”)
Exhibit 1008	Mallinali Ramirez-Corona et al., <i>Multi-label Classification for Tree and Directed Acyclic Graphs Hierarchies</i> , Probabilistic Graphical Models, PGM 2014, (2014) (“Ramírez-Corona”)
Exhibit 1009	Nitesh V. Chawla et al., <i>SMOTE: Synthetic Minority Over-sampling Technique</i> , Journal of Artificial Intelligence Research 16 (2002) 321-357 (“Chawla”)
Exhibit 1010	Tong Xiao et al., <i>KNN and Re-ranking Models for English Patent Mining at NTCIR-7</i> , Proceedings of NTCIR-7 Workshop Meeting, (2008) (“Xiao”)
Exhibit 1011	U.S. Patent No. 7,028,250 to Ukrainczyk et al. (“Ukrainczyk”)
Exhibit 1012	Eric Nunes et al., <i>Darknet and Deepnet Mining for Proactive Cybersecurity Threat Intelligence</i> , Proceedings of the 2016 IEEE Conference on Intelligence and Security Informatics (“Nunes”)
Exhibit 1013	<i>Top 10 Dark Web and Deep Web Forums of 2025</i> , https://cyble.com/knowledge-hub/top-10-dark-web-forums/ , (“Cyble”)
Exhibit 1014	<i>Dark Web vs. Deep Web: 5 Key Differences</i> , https://www.Spiceworks.com/it-security/security-general/articles/dark-web-vs-deep-web/ , (“Spiceworks”)
Exhibit 1015	Thomas Hoffman et al., <i>Hierarchical Document Categorization with Support Vector Machines</i> , CIKM ’04: Proceedings of the thirteenth ACM international conference on Information and knowledge management, (2004) (“Hoffman”)
Exhibit 1016	Isuf Deliu, Carl Leichter, and Katrin Franke, <i>Extracting cyber threat intelligence from hacker forums: Support vector machines</i>

	<i>versus convolutional neural networks</i> , 2017 IEEE International Conference on Big Data (BIGDATA), 3648-3656 (2017) (“ <i>Deliu Paper</i> ”)
Exhibit 1017	Xinqi Zhu <i>et al.</i> , <i>B-CNN: Branch Convolutional Neural Network for Hierarchical Classification</i> , <i>arXiv.org</i> , (2017) (“ <i>Zhu</i> ”)
Exhibit 1018	Zhicheng Yan <i>et al.</i> , <i>HD-CNN: Hierarchical Deep Convolutional Neural Networks for Large Scale Visual Recognition</i> , 2015 IEEE International Conference on Computer Vision (ICCV), (2015) (“ <i>Yan</i> ”)
Exhibit 1019	Quoc Le <i>et al.</i> , <i>Distributed Representations of Sentences and Documents</i> , Proceedings of the 31st International Conference on Machine Learning, PMLR 32(2):1188-1196, (2014) (“ <i>Le</i> ”)
Exhibit 1020	Revanth Patil <i>et al.</i> , <i>Multi-class and Multi-label classification of Darkweb Data</i> , Arizona State University, (Approved March 2018) (“ <i>Patil Thesis</i> ”)
Exhibit 1021	Alex Smola <i>et al.</i> , <i>Introduction to Machine Learning</i> , (2008) (“ <i>Smola</i> ”)
Exhibit 1022	Okapi-BM25, https://en.wikipedia.org/wiki/Okapi_BM25 , (“ <i>Okapi-BM25 Wikipedia Webpage</i> ”)
Exhibit 1023	M. Hassler <i>et al.</i> , <i>Text preparation through extended tokenization</i> , Data Mining VII: Data, Text and Web Mining and their Business Applications, (2006) (“ <i>Hassler</i> ”)
Exhibit 1024	Jasmeet Singh <i>et al.</i> , <i>A systematic review of text stemming techniques</i> , Artificial Intelligence Review, (2016) (“ <i>Singh</i> ”)
Exhibit 1025	Tianjun Fu <i>et al.</i> , <i>A Focused Crawler for Dark Web Forums</i> , Journal of the American Society for Information Science and Technology, (2010) (“ <i>Fu</i> ”)
Exhibit 1026	Mitch Macdonald <i>et al.</i> , <i>Identifying Digital Threats in a Hacker Web Forum</i> , IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, (2015) (“ <i>Macdonald</i> ”)
Exhibit 1027	Roman Eisner <i>et al.</i> , <i>Improving Protein Function Prediction using the Hierarchical Structure of the Gene Ontology</i> , 2005 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology, (2005) (“ <i>Eisner</i> ”)
Exhibit 1028	Alaa Alahmadi, <i>Automatic text classification using bag of words and bag of concepts based representations</i> , University of Limerick, (2016) (“ <i>Alahmadi</i> ”)
Exhibit 1029	Microsoft Computer Dictionary, 5 th ed., 2002, excerpts (“ <i>Microsoft Computer Dictionary</i> ”)

Exhibit 1030	Curriculum vitae (CV) of Dr. Erez Zadok.
---------------------	--

CERTIFICATION OF WORD COUNT

The undersigned certifies pursuant to 37 C.F.R. §42.24 that the foregoing Petition for *Inter Partes* Review, excluding any table of contents, mandatory notices under 37 C.F.R. §42.8, certificates of service or word count, or appendix of exhibits, contains 12,498 words according to the word-processing program used to prepare this document (Microsoft Word).

Dated: August 25, 2025

By: /s/ Paul R. Hart
Paul R. Hart, Reg. No. 59,646

COUNSEL FOR PETITIONER

**CERTIFICATE OF SERVICE ON PATENT OWNER
UNDER 37 C.F.R. § 42.105**

Pursuant to 37 C.F.R. §§ 42.6(e) and 42.105(b), the undersigned certifies that on August 25, 2025, a complete and entire copy of this Petition for *Inter Partes* Review and Exhibits were provided via Federal Express to the Patent Owner by serving the correspondence address of record for the '900 Patent:

POLSINELLI PC
PO Box 140310
Kansas City, MO 64114-0310

Further, a courtesy copy of this Petition for *Inter Partes* Review was sent via email to Patent Owner's litigation counsel:

Jay P. Kesan (jay.kesan@kkd-law.com)
Cecil E. Key (cecil.key@kkd-law.com)
Alfonso Garcia Chan (achan@cahill.com)
Britton F. Davis (bdavis@cahill.com)
Christopher C. Campbell (ccampbell@cahill.com)
Halima Shukri Ndai (hndai@cahill.com)

Respectfully submitted,

By: /s/ Paul R. Hart
Paul R. Hart, Reg. No. 59,646

COUNSEL FOR PETITIONER