Evaluation Details of Biocreative BEL Task 1

Contents

General Evaluation Criteria

- BEL statements must be syntactically correct to be accepted for evaluation: BEL terms (representing entities) must be complete and in a correct format (containing either an abundance or a process function depending on the namespace), otherwise a submission will not be evaluated

- BEL statements can be submitted as full BEL statements or as fragments of full BEL statements. A submitted BEL statement is automatically cut into its fragments to ensure evaluations on different levels (see below).

- **There will be two submission deadlines (see Important Dates): After the first submission deadline*, gold standard entities will be released. These can be used by the participants to prepare a second submission. A unlimited number of BEL statements/BEL fragments will be accepted in each run. F-Score will be the primary evaluation metric and therefore high recall approaches will be penalized.*

The task is evaluated in several sub-levels, as described below. This method of evaluation is intended to enable partial participation. Even if not all levels are fulfilled, a reasonable evaluation score can be achieved. For example, teams that do not want to produce normalized entities (see "Evaluation on Term-Level"), can use a placeholder instead as will be explained in the following.

NEW:

- ****NEW DATES** (See Important Dates)**

- **an overview of concrete examples of how submitted BEL statements are scored on different levels can be found here: Example Submission .**

- **an evaluation interface where participants can test generated statements is available under http://bio-eval.scai.fraunhofer.de/cgi-bin/General_server.rc**
    - **documentation of the interface can be found here: Evaluation Framework**

Evaluation on different Levels

During evaluation, we will split the BEL statements produced by the participants and evaluate them on the following levels:

- Term-level

- Function-level

- Relationship-level

- Full statement evaluation

- Overall evaluation

Evaluation Example – full BEL statement and the parts evaluated on the different levels:

BEL Statement	p(HGNC:BCL2A1) decreases bp(GOBP:"apoptotic process")	act(p(MGI:Hras)) increases p(MGI:Mmp9)
Evidence Sentence	*We demonstrate that the Bfl-1 protein suppresses apoptosis induced by the p53 tumor suppressor protein in a manner similar to other Bcl-2 family members such as Bcl-2, Bcl-xL and EBV-BHRF1.*	*Cells with activated ras demonstrated high level of expression of 72-kDa metalloproteinase (MMP-2, gelatinase A), and 92-kDa metalloproteinase (MMP-9, gelatinase B) compared with cells containing SV40 large T antigen alone.*

Term-level Evaluation	`p(HGNC:BCL2A1)`
	
	 `bp(GOBP:"apoptotic process")`
		`p(MGI:Hras)`
		
		 `p(MGI:Mmp9)`

Function-level Evaluation	–	`act(p(MGI:Hras))`
Relationship-level Evaluation	`p(HGNC:BCL2A1) decreases bp(GOBP:"apoptotic process")`	`p(MGI:Hras) increases p(MGI:Mmp9)`
Full-statement evaluation	`p(HGNC:BCL2A1) decreases bp(GOBP:"apoptotic process")`	`act(p(MGI:Hras)) increases p(MGI:Mmp9)`

NOTE:

- BEL functions are evaluated on the Term level (Term Functions: Abundance/Process Functions) and the Function Level (Other Functions). A full overview of all BEL functions and how they are considered in the evaluation can be found in *All Functions Evaluation Overview*.

- **For each evaluation level, precision, recall and F-score will be calculated**

Evaluation on Term-Level

On Term-level, the correctness of BEL terms will be evaluated.

BEL terms are built from entities, their namespaces and associated abundance or process functions.

The evaluation of BEL terms includes the following:

- Correctness of the discovered entities
- Correctness of associated namespaces and their format
- Correctness of associated abundance/process functions (note: if these are not provided, a term is not accepted for evaluation)

BEL Terms are evaluated at one sub-levels:

1) **"Term Level":** Correctness of complete BEL term; Are the abundance/process functions as well as the namespaces and identifiers correct?

NOTE:

- We do not expect organism disambiguation. Entities are accepted based on HGNC, MGI or EGID if they are equivalent.

- **On the term level, instead of exact namespaces and identifiers, we accept placeholders of the format "PH: placeholder" for entities. Abundance or process functions are still necessary to ensure correct BEL syntax. However, using the same abundance function with placeholders only affects scores on the T_s level.**

  Placeholder Argument Example:

  - p(PH:placeholder) : any undefined entity

  - p(PH:placeholder): an undefined gene or protein

  - a(PH:placeholder): an undefined chemical

  - path(PH:placeholder): an undefined disease

- bp(PH:placeholder): an undefined biological process

- **Placeholders can either be used as a standard solution (evaluated on all levels where correct arguments are not taken into account) or as a last resort to generate a BEL term for entities where a normalization through namespace and identifier is not possible.**
- *Placeholders never score a FP (false positive) as incorrect BEL terms would do. They are only counted as FNs (false negatives) and therefore only influence recall but not precision values.*

BEL Term Format	Associated Term Function			Associated Namespaces	Other acceptable functions	BEL term Example	Comments
	Short Function Name	Long Function Name	Function Type				
p(Namespace:Entity)	`p()`	proteinAbundance()	abundance	HGNC, MGI, EGID	`g(), r(), m()`	`p(HGNC:MAPK14)`	*We recommend using the p() function only. The p() function will be accepted in place of all other possible functions (i.e. the functions g(), r() and m()).*
bp(Namespace:Entity)	`bp()`	biologicalProcess()	process	GOBP	–	`bp(GOBP:"cell migration")`	
path(Namespace:Entity)	`path()`	pathology()	process	MESHD	–	`path(MESHD:Fibrosis)`	
a(Namespace:Entity)	`a()`	abundance()	abundance	CHEBI	–	`a(CHEBI:dioxidaniumyl)`	

Evaluation on Function-Level

On function-level, the correctness of discovered function will be evaluated. Functions are only accepted together with their argument BEL terms.

Functions are evaluated at two sub-levels:

1) **"Function Level":** Correctness of functions together with their arguments; Is a function associated to the correct BEL terms?

2) **Secondary Function Level":** Correctness of a function only, regardless of the correctness of their term-arguments (*BUT: BEL terms or placeholders need to be present!*)

For each of these sub-levels, an F-Score will be calculated.

NOTE:

- *Even though only the functions listed above will be assessed in the function-level evaluation, the BEL statements are expected to be syntactically correct and complete BEL terms (containing an abundance or process function together with a namespace and identifier or a placeholder) are expected.*

- *On the function level, instead of exact namespaces and identifiers, we accept placeholders of the format "PH: placeholder" for function term-arguments. Abundance or process functions are still necessary to ensure correct BEL syntax. However, the same abundance function can be used for all namespaces.*

    Placeholder Argument Example:

    - act(p(PH:placeholder)) : an undefined entity

    - **Placeholders can either be used as a standard solution (evaluated on all levels where correct arguments are not taken into account) or as a last resort to generate a BEL term for entities where a normalization through namespace and identifier is not possible.**
    - *Placeholders never score a FP (false positive) as incorrect BEL terms would do. They are only counted as FNs (false negatives) and therefore only influence recall but not precision values.*

- **The arguments of the complex() function are scored as correct if at least one argument-term is correct**

The following functions are present in the test set and will be assessed:

Function Type	Function	Example	Comments
Abundances	`complex()`	complex(p(HGNC:IL10),p(HGNC:TGFB1),p(HGNC:FASLG))	*Full credit is given if at least one argument of the complex() function is correct.*
Transformations	`tloc()`	Original statement: deg(p(MGI:Ctnnb1)) decreases tloc(p(MGI:Ctnnb1),GOCCID:0005737,GOCCID:0005634) Also Accepted : deg(p(MGI:Ctnnb1)) decreases tloc(p(MGI:Ctnnb1)) Accepted Statement with Placeholders (only scores at sub-level 1): deg(p(PH:placeholder)) decreases tloc(p(PH:placeholder))	*Location arguments are not expected but will be accepted if syntactically correct. No credit will be given for location arguments.* The tloc() function will be accepted and credited in place of the following functions: sec(), surf()
	`deg()`	act(p(HGNC:MMP1)) increases deg(p(HGNC:COL2A1))	
Modifications	`pmod()`	Original Statement: bp(GOBP:"response to tumor cell") increases p(MGI:Bad,pmod(P,S,112)) Also Accepted: bp(GOBP:"response to tumor cell") increases p(MGI:Bad,pmod(P)) Accepted Statement with Placeholders (only scores at sub-level 1): p(PH:placeholder) increases p(PH:placeholder,pmod(P))	*Only pmod(P) will be evaluated. Additional arguments are not expected but will be accepted if syntactically correct. No credit will be given for additional arguments.* The pmod() function will not be accepted and credited in place of any other functions:
Activities	`act()`	Original Statement (no credit will be given for kin() function): kin(p(MGI:Kdr)) increases p(MGI:Pecam1) Accepted Statement: act(p(MGI:Kdr)) increases p(MGI:Pecam1) Accepted Statement with Placeholders (only scores at sub-level 1): act(p(PH:placeholder)) increases p(PH:placeholder)	*Credit will only be given for the act() function only. The act() function will be accepted and credited in place of any other activity function listed.* The act() function will be accepted and credited in place of the following functions: cat(), chap(), gtp(), kin(), pep(), phos(), ribo(), tscript(), tport() *These functions are not accepted for evaluation; It is necessary to use act() instead.*

Evaluation on Relationship-Level

On relationship level, the relationship contained in a BEL statement will be evaluated. BEL relationships have a subject, predicate, object structure. The subject and the object are the entities involved in a relationship in the format of BEL terms, the predicate is the relationship between them. Only these components of a BEL statement will be taken into account for the evaluation on relationship-level. Other functions which are not term functions, will be ignored.

Relationships will be evaluated on two levels:

1) **"Relationship Level":** Full Relationships; subject and object need to be correct.

2) **"Secondary Relationship Level":** Partial Relationships; relationships containing two correct units, either

- a correct relationship together with an incorrect subject and a correct object
- a correct relationship together with a correct subject and an incorrect object
- an incorrect relationship together with a correct subject and a correct object

For both levels, an F-score will be calculated.

NOTE:

- Only the entities and the relationships are considered in the relationship-level evaluation. The correctness of functions (except for the complex() function) that are part of a relationship is not taken into account on this leval. However, complete BEL terms (containing an abundance or process function) are expected

- The presence of the complex() function is evaluated on the relationship level. A complex function is evaluated as correct if at least one of its term-arguments is correct.

- Also on relationship level we do not expect organism disambiguation. Entities are accepted based on HGNC, MGI or EGID if they are equivalent.

- *Instead of exact namespaces and identifiers, we accept placeholders of the format "PH: placeholder" (see term and function level for examples). Placeholders are only scored as FN (false negatives) but not as FP (false positives).*

- *On the relationship level, we also accept a placeholder "association", which can be used if the relationship type is not known. The placeholder "association" only scores as FN (false negative) but not as FP (false positive) on the relationship level and the full statement level.*

The following relationships will be assessed:

Relationship	Alternative Symbol	Example: Statement from Gold Standard	Expected/Evaluated on Relationship Level: only the relationship and its term-arguments	Examples for Partial Relationships/Use of Placeholders (only scoring at secondary relationship level evaluation)	Comments

`decreases`	`-|`	`kin(p(MGI:Kdr))`
		`decreases`
		`p(MGI:Tek)`
			`p(MGI:Kdr)`
			`decreases`
			`p(MGI:Tek)`
				Partial Relationship:
				
				`p(MGI:Kdr) decreases p(MGI:wrong identifier)`
				
				Entity Placeholder:
				
				`p(PH:Placeholder) decreases p(MGI:Tek)`
				
				Association Placeholder:
				
				`p(MGI:Kdr) association p(MGI:Tek)`	

`directlyDecreases`	`=|`	Accepted Statement:
		
		`act(p(HGNC:MDM2)) directlyDecreases p(HGNC:TP53)`
		
		Also Accepted:
		
		`act(p(HGNC:MDM2)) decreases p(HGNC:TP53)`
			`p(HGNC:MDM2) directlyDecreases p(HGNC:TP53)`
			
			`p(HGNC:MDM2) decreases p(HGNC:TP53)`
			
			(either version is accepted and scores the same credit)
				Partial Relationship:
				
				`p(HGNC:wrong identifier) directlyDecreases p(HGNC:TP53)`
				
				Entity Placeholder:
				
				`p(HGNC:MDM2) decreases p(PH:Placeholder)`
				
				Association Placeholder:
				
				`p(HGNC:MDM2) association p(HGNC:TP53)`
					decreases will be accepted in place of directlyDecreases with the same credit given

increases	->	p(HGNC:ARRB1) increases kin(p(HGNC:MAPK1))	p(HGNC:ARRB1) increases p(HGNC:MAPK1)	Partial Relationship: p(HGNC:wrong identifier) increases p(HGNC:MAPK1) Placeholder: p(PH:Placeholder) increases p(HGNC:MAPK1) Association Placeholder: p(HGNC:ARRB1) association p(HGNC:MAPK1)	
directlyIncreases	=>	Accepted Statement: p(HGNC:VEGFB) directlyIncreases act(p(HGNC:FLT1)) Also Accepted: p(HGNC:VEGFB) increases act(p(HGNC:FLT1))	p(HGNC:VEGFB) directlyIncreases p(HGNC:FLT1) p(HGNC:VEGFB) increases p(HGNC:FLT1) (either version is accepted and scores the same credit)	Partial Relationship: p(wrong namespace:wrong identifier) directlyIncreases p(HGNC:FLT1) Placeholder: p(HGNC:VEGFB) increases p(PH:Placeholder) Association Placeholder: p(HGNC:VEGFB) association p(HGNC:FLT1)	*increases* will be accepted in place of directlyIncreases with the same credit given

Special Case	Statement from Gold Standard	Expected/Evaluated on Relationship Level: only the relationship and its term-arguments	Comment
Example of a Relationship with complex as an argument	cat(complex(p(HGNC:CREBBP),p(HGNC:EP300))) increases p(HGNC:KLF1,pmod(A))	complex(p(HGNC:CREBBP),p(HGNC:EP300))) increases p(HGNC:KLF1) complex(p(HGNC:CREBBP)) increases p(HGNC:KLF1) complex(p(HGNC:EP300)) increases p(HGNC:KLF1) complex(p(HGNC:CREBBP),p(PH:Placeholder))) increases p(HGNC:KLF1) (all above versions are accepted and score the same credit)	*The complex() function is the only function that is evaluated on the function as well as on relationship level (not considering process and abundance functions but only other functions).* *A complex function is evaluated as correct if at least one of its term-arguments is correct.*

Full Statement

Evaluates if a full BEL statement is correct and complete.

NOTE:

- *Submission of fragments of BEL statements can score higher in other levels but will damage the full statement level.*

- *Instead of exact namespaces and identifiers, we accept placeholders of the format "PH: placeholder" also in full statements (see term and function level for examples). If a full statement is correct but BEL-term (representing entities) are expressed as placeholders instead of namespaces and identiifers, only a FN (false negative) but no FP (false positive) is counted.*

Overall Evaluation

A final overall score will be calculated from the results of all evaluation levels. Full discovered statements will be scored by their amount of coverage compared to the gold standard statement.