Generating a Logical Structure for Virtualizing Physiotherapy Instructions through NLP

Sandeep Kumar Dash[1], Partha Pakray[2]
National Institute of Technology Mizoram
India
sandeepdashcet@gmail.com
parthapakray@gmail.com

Alexander Gelbukh[3]
CIC, IPN Mexico
India
gelbukh@gelbukh.com

ABSTRACT: *This paper describes a proposed framework for virtualizing the documented physiotherapy instructions. It tries to bridge the gap between human understanding and the written manuals of instructions for physiotherapy through text processing. A structure has been developed that reflects the requirement for processing natural language text to virtual action. Techniques of Natural Language Processing involving semantic and spatial information processing carries importance in this approach.*

Keywords: Text Processing; Virtual Action

Received: 1 June 2016, Revised 3 July 2016, Accepted 9 July 2016

1. Introduction

General human tendency is to visualize a scenario while going through any document or text. So converting input natural language text into virtual action holds much importance with respect to the complexity faced mainly in understanding of written manuals, simulation of production line, story description for students as an aid in literacy and in other such applications. Natural Language text made to describe a scene is much easier than complex manipulation interfaces through which objects are constructed and precisely positioned within scenes. The complexity lies with natural language in which both linguistic and real-world knowledge, in particular knowledge about the spatial and functional properties of objects; prepositions and the spatial relations they convey is often ambiguous. Also verbs and how they resolve to poses and other spatial relations [1] as well needs to be analyzed in the input text for a proper scene building. Furthermore, implicit facts about real world has also to be analyzed from the text which is rarely mentioned therewith.

Physiotherapy is one such area where textual descriptions of the exercises are difficult to interpret. The various body parts, joints and their angle along with direction of movement carries utmost importance while carrying out specific physiotherapies. This is one of the challenging area where interpretation of textual content has to be perfect or else consequences can be damaging for health.

In this paper we are introducing structure for representing text describing different body part movements during physiotherapy exercises. We will describe how the input textual description of a physiotherapy exercise can be converted into a virtual action describing the same. The method is proposed for some of the therapies only, which can be extended for others after further research.

2. Related Work

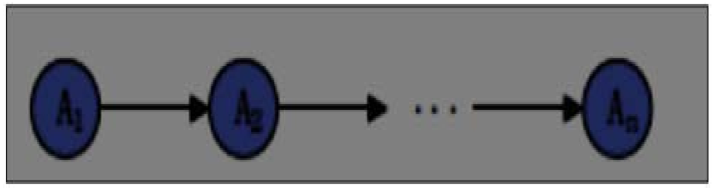
There has been a great number of work which implements natural language text as input for generating 3D scene as a virtual scenario. Researchers have utilized various text processing methods in order to build structures which can be made to generate the scene. Mainly these systems follow basic four steps. First is to parse the input text into a structure or template establishing spatial relation between the different objects mentioned in it. Secondly using real world knowledge, find any implicit constraints also which is not mentioned in the text. In third step the scene templates are converted into geometric 3D scene. Lastly optimize the placement of objects as per the templates. These types of work are influencing in our research in the sense that the spatial information about the objects mentioned in the sentence carries importance, as it can also be different body parts placement during a physiotherapy along with the items used during this as balls, chairs, wall, floor, sitting mat etc.

The pioneering work in this area was carried out by [1]. They developed an automatic text-to-scene conversion system named '*WordsEye*' [1]. Their first system was designed with a large library of 3D objects to depict scenes from input text, whereas their current system contains 2,200 3D objects and 10,000 images and a lexicon of approximately 15,000 nouns. It supports language-based control of objects, spatial relations, and surface properties (e.g., textures and colors); and it handles simple Co-reference resolution, allowing for a variety of ways of referring to objects [1]. They have utilized a self-designed lexical knowledge-base, called the SBLR (Scenario-Based Lexical Resource) which consists of an ontology and lexical semantic information extracted from WordNet [2] and FrameNet [3]. Their system works by first tagging and parsing the input text, using part of speech tagger [4] and parser [5]. The parser output is then converted into a dependency structure which is processed to resolve anaphora and other Co-references. The dependency structure is utilized to create semantic nodes from which the semantic relations of lexical items are formed with the help of information in the SBLR. These semantic relations are then converted to a final set of graphical constraints representing the position, orientation, size, color, texture, and poses of objects in the scene. This final structure was then utilized to create the 3D scene with the help of a graphics rendering software.

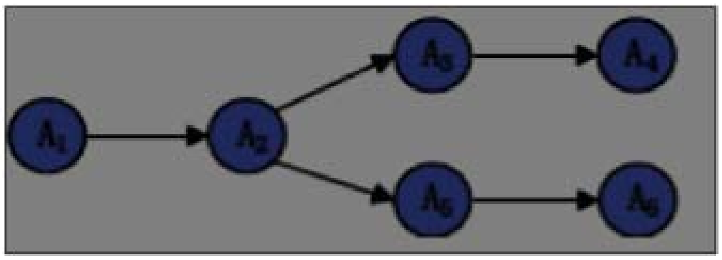
Research by [6] has shown an approach of finding the implicit spatial relationship between objects described in a scene. They considered the output scene to be generated as a graph whose nodes are objects mentioned either implicitly or explicitly in the text and the edges as the semantic relationships among them. The semantics of a scene were described using a 'scene template' and the geometric properties using a 'geometric scene'. In their approach scene template is a triplet T = (O, C, Cs) which consists of object descriptions O, constraints C on the relationships between the objects and a scene type Cs. Here object description provides information about the object as category label, basic attributes such as color and material, and number of occurrences in the scene. Spatial relations between objects were expressed as predicates of the form supported_by(oi , oj) or left(oi , oj) where oi and oj are recognized objects. Geometric Scene represents concrete geometric representation of a scene. It consists of a set of 3D model instances – one for each object – that capture the appearance of the object. To exactly position the object they also derived a transformation matrix that represents the position, orientation, and scaling of the object in a scene. A scene template was generated by selecting appropriate models from a 3D model database and determining transformations that optimize their layout to satisfy spatial constraints.

The system [7] has created a visual action design environment that improves action design for intelligent toy robot. The environment automatically creates action sequence files in terms of txt or xml from given action sequences. Action sequences are nothing but set of aggregated arrays of actions as per their time of occurrences. These actions may be parallel actions with complete or in-part synchronous with time, linear action or repeated serial actions. They can be better understood by Fig. 1.

These action sequences can be created in the visualized design environment and the action sequence file can be automatically generated thereafter. These files contain action sequences in terms of tags and data blocks as shown in Table 1.



(a)



(b)

Figure 1. Types of action sequences (a) Linear Action (b) Parallel Action

Begin	Action
Begin	...
LinearAction	End Action
Action	End LinearAction
…(action	Begin
information)	LinearAction
End Action	Action
Action	…
…	End Action
End Action	Action
Begin	…
Parallel Action	End Action
Begin	End LinearAction
LinearAction	End ParallelAction
Action	End LinearAction
	End

Table 1. Tags And Data Blocks For Action Sequences (Source: [7])

In above the tag “Begin LinearAction/ End LinearAction” represents linear action, tag “Begin ParallelAction/ End ParallelAction” represents parallel action and the tag “Action/End Action” represent data block which represents information about a particular action [7]. The intelligent toy robots are limited to only rotational freedom. Hence any action is defined logically as Ai (Joint, Angle, Axis, Absolute, Time, StartTime, Isparallel). The parameters respectively represent joint that exerts action, angle of rotation, axis of rotation, absolute or relative value of angle, and timeline for adjusting & harmonizing the coherent action sequence, order of size of action and finally whether the node is an action node or substitute node. Further they have used the queue data structure to store the action files as per their order of StartTime. Finally virtual action was generated by matching action sequence files into corresponding joints of intelligent toy robot model. They used Eon studio for virtual simulation purpose.

In another innovative research funded by EU named MUSE (Machine Understanding for interactive Storytelling) [8] project, the aim is to bring texts to life by developing an innovative text-to-virtual-world translation system. It mainly evaluates two scenarios: story-telling and patient education materials. The idea involves semantic role recognition of a sentence, spatial relations between objects and chronology of events.

3. Dataset Description

The dataset contains textual manuals for different types of physiotherapy instructions involving various body parts as Hand, Foot, Neck, Back, Shoulder, Knee etc. For example knee may involve exercises as 'Knee squats', 'Leg cross', 'Leg stretch', 'Sitstands', 'Step ups' etc each maintained as different text manuals. Consider the following instruction consisting of five sentences.

Instruction 1: *[Name: Leg stretch]* Sit on the floor with your legs stretched out in front. Keeping your foot to the floor, slowly bend one knee until you feel it being comfortably stretched. Hold for 5 seconds. Straighten your leg as far as you can and hold for 5 seconds. Repeat 10 times with each leg.

The Dataset contains various such instructions involving different body parts, some of which are mentioned in Table 2. For training the system 70% of instructions will be used (approximately 200 instructions) and 30% will be used as test set.

Body Part	**Instructions**
Knee	Leg stretch Leg cross Knee squats Sitstands Step ups
Neck	Neck stretch Neck Tilt Neck Turn
Shoulder	Door Lean Door Press Pendulum Exercise Shoulder Stretch
Foot	Achilles tendon and plantar Plantar fascia stretch Sitting plantar fascia stretch Towel pickup
Back	Back stretch Deep Lunge Knees to chest One-leg stand (front) Pelvic tilt

Table 2. Dataset of Physiotherapy Instructions

4. Proposed Framework

The framework involves natural language processing techniques, which can be described as below;

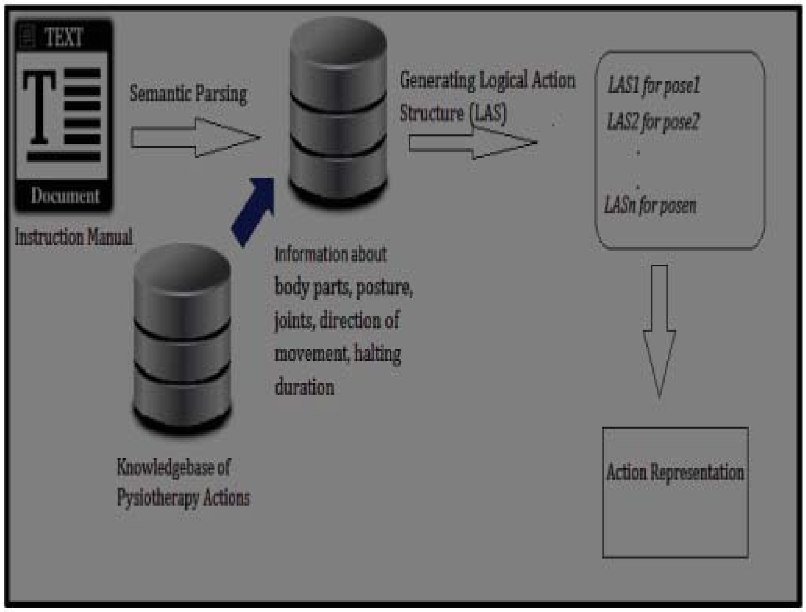


Figure 2. Proposed framework of Logical Action Representation

In the proposed framework the knowledge base to be used will contain information about type of posture each therapy generate by involving the joints of the respective parts, the prescribed angle between joints, type of movements and the necessary time duration of halting in the pose as well as the number of times the action is to be repeated. This will help to have the implicit information produced for a better action representation. In the initial stage we would incorporate the following information in our system.

4.1 Type of Body Planes

For describing the structural positions and directions of functional movement of the body the standard posture is body facing forward, the hands at the sides of the body, with the palms facing forward, and the feet pointing straight ahead. Body planes are derived from dimensions in space and are oriented at right angles to one another. The median sagittal plane, also called the midsagittal plane, divides the body into right and left halves. The coronal plane is vertical and extends from side to side. The transverse plane is a horizontal plane and divides a structure into upper and lower components. We will use these three types of planes for representing different postures.

4.2 Type of Body Axis

Axis represents a line around which movement will occur. We will consider three types of movements possible around each joint rotation, translation, and curvilinear motion. All movements that occur about an axis are considered rotational, whereas linear movements along an axis and through a plane are called translational. Curvilinear motion occurs when a translational movement accompanies rotational movements.

4.3 Type of Joints

The three types of joints we will consider are mainly based upon how much movement they allow. A synarthrosis joint permits no movement as the gomphoses that connect the teeth to the skull. An amphiarthrosis joint allows a slight amount of movement at the joint. For ex. pubic symphysis of the hips. The third type is the diarthrosis joint. Diarthroses have the highest range of motion of any joint and include the elbow, knee, shoulder, and wrist.

Some of the specific movements we will consider around joints are in Table 3:

Types	Description
Flexion	Bending parts at a joint so that the angle between them decreases and the parts come closer together (bending the lower limb at the knee).
Extension	Straightening parts at a joint so that the angle between them increases and the parts move farther apart (straightening the lower limb at the knee).
Hyperextension	Excess extension of the parts at a joint, beyond the anatomical position (bending the head back beyond the upright position).
Eversion	Turning the foot so the sole faces laterally.
Inversion	Turning the foot so the sole faces medially.
Protraction	Moving a part forward (thrusting the chin forward).
Retraction	Moving a part backward (pulling the chin backwards).
Elevation	Raising a part (shrugging the shoulders).
Depression	Lowering a part (drooping the shoulders).

Table 3. Various Movements

We are introducing LAS (Logical Action Structure) that will describe logically each pose for a particular therapy. It will involve information about body plane, axes, type of movement mentioned in the described text. These details will be used to form the action representation for each pose, thereby helping to form action sequences. LAS is having the following format.

{b_plane, body parts, axes, m_type, duration_of_halt_in_pose}
where:

b_plane	:	Body Plane
axes	:	Axes of movements
m_type	:	Type of Movement
duration_of_halt_in_pose	:	This will provide the time gap between between consecutive actions

Through efficient text processing methods along with implicit information incorporation from our developed knowledgebase of different therapy poses, each structure can represent its logical action which can lead sequentially to generate the exact virtual action prescribed in the written manuals.

References

[1] Coyne, B., Sproat, R., Hirschberg, J.: Spatial relations in text-to- scene conversion. *In*: Computational Models of Spatial Language Interpretation, Workshop at Spatial Cognition (2010)

[2] Fellbaum, C. (1998). WordNet: an electronic lexical database. MIT Press (1998)

[3] Baker, C., Fillmore, C., Lowe, J.: The Berkeley FrameNet Project. COLING-ACL (1998)

[4] Church, K. W. (1998). A stochastic parts program and noun phrase parser for unrestricted text. *In*: Proceedings of the Second Conference on Applied Natural Language Processing (1988) 136-143

[5] Collins, Michael. et al. (1999). A statistical parser for Czech, *In*: Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics. Association for Computational Linguistics (1999)

[6] Chang, Angel X., S, Manolis., Manning, CD. (2014). Learning Spatial Knowledge for Text to 3D Scene Generation. EMNLP (2014) 2028-2038.

[7] Mo, Jianqing., Hanwu, He., Zhang, H. (2012). Virtual simulation of intelligent toy robot driven by the action sequence files.

Computer Science and Automation Engineering (CSAE), 2012 IEEE International Conference on. V. 1. IEEE (2012) 401-404.

[8] De, Mulder W., Do, Thi NQ., van den, Broek P., Moens, MF. (2013). Machine understanding for interactive storytelling. *In:* Proceedings of KICSS 2013- 8th international conference on knowledge, information, and creativity support systems (2013) 73-80.