

А.Г. Додонов, Д.В. Ландэ
Украина, Киев, ИПРИ НАН Украины
dodonov@ipri.kiev.ua, dvl@visti.net

ЖИВУЧЕСТЬ ИНФОРМАЦИОННЫХ СЮЖЕТОВ

Рассматриваются вопросы живучести информационных сюжетов в информационном пространстве с точки зрения теории систем. Приведены примеры, алгоритмы и модели формирования информационных сюжетов, рассмотрены подходы к количественной оценке их живучести.

Ключевые слова: инфосюжет, система, живучесть, информационное пространство, информационная безопасность

Определим информационный сюжет (далее – инфосюжет) как логически связанное множество информационных сообщений, посвященных одной тематике или одному событию. Инфосюжеты являются своеобразными кластерами в информационном пространстве, которые, как будет показано ниже, полностью удовлетворяют общему определению систем.

Живучесть инфосюжета может рассматриваться и как его объективное свойство, которое зависит от тематики, аудитории, времени и т.д., и как качество, которое хотят придать ему в случае искусственного формирования, например, при проведении информационных операций.

Живучесть – это фундаментальное свойство сложных систем. Многие системы изначально обладают свойством живучести, что позволяет им сохранять целостность, выполнять свои функции и развиваться независимо от наличия неблагоприятных воздействий со стороны внешней среды и внутренней деградации [1]. Живучие системы способны поддерживать непрерывное выполнение своих основных функций (в случае инфосюжетов – информировать о различных аспектах тематики или события), временно или постоянно отказываясь от выполнения менее важных, изменять свою структуру и поведение, находить и выполнять новые функции, необходимые для успешного противостояния внешней среде (информационному пространству), приспосабливаясь к условиям функционирования [2].

Система в общем смысле – это совокупность сильно связанных объектов, обладающая свойствами организации, связности, целостности и членимости. Соответственно, инфосюжет можно рассматривать как контентную систему, совокупность информационных сообщений, взаимосвязанных взаимными ссылками, цитированием, гиперссылками, общей лексикой и т.д., выделенных из среды (информационного пространства) на определенное время (время актуальности) с определенной целью. В самом деле, инфосюжет – это совокупность сильно связанных объектов, обладающая свойствами организации, связности, целостности (определяемой тематикой или событием) и членимости (на отдельные сообщения). Известно, что системы в общем случае могут разделяться на два класса: целенаправленные и нецеленаправленные. Инфосюжеты как системы могут относиться как к первому, так и ко второму классу. При этом целенаправленные сюжеты могут порождаться в ходе активных рекламных кампаний, в качестве отвлекающих факторов, дезинформирования и других процессов [3]. Вследствие этого живучесть информационных сюжетов имеет важнейшее значение для информационной безопасности.

Рассмотрим свойства инфосюжетов, присущие им как системам:

- синергетичность – в случае инфосюжетов направленность отдельных информационных сообщений усиливает информационную функцию всего инфосюжета;
- приоритет интересов инфосюжета перед интересами его компонентами (отдельными сообщениями);
- эмерджентность – информационные цели (функции) компонент не всегда совпадают с целями (функциями) всего инфосюжета, свойства инфосюжета не сводятся к сумме свойств составляющих его сообщений;
- целостность – первичность целого инфосюжета по отношению к отдельным информационным сообщениям;
- структурность – возможна декомпозиция инфосюжета на компоненты (сообщения), установление связей между ними;
- коммуникативность – существование сложной системы коммуникаций инфосюжетов с информационным пространством, взаимодействие и инфосюжета и информационного пространства;
- адаптивность – возможность адаптации параметров инфосюжета к изменяющимся параметрам внешней среды;
- надежность – существование инфосюжета при выходе из строя отдельных его компонент (информационных сообщений), сохраняемость значений параметров системы в течение определенного периода.

Существует еще ряд системных свойств инфосюжетов, таких как: интегративность – наличие системообразующих, системосохраняющих факторов; эквивинальность – способность инфосюжетов достигать состояний независимых от исходных условий и определяющихся только параметрами системы; наследственность; самоорганизация и т.д.

Как на весь инфосюжет, так и на отдельные информационные сообщения, входящие в него могут оказываться различные дестабилизирующие информационные воздействия, например, удаление отдельных статей с веб-сайтов, уничтожение или отключение информационных серверов, публикация сообщений, которые искажают исходный инфосюжет или порождение нового инфосюжета, который может понизить актуальность исходного инфосюжета. При этом нарушение целостности инфосюжета на фоне снижения актуальности его компонент влечет за собой дезорганизацию, понижение живучести и нарушение целостности, т.е. потерю важнейших функций инфосюжетов [4].

Информационные сюжеты могут быть представлены как сетевые структуры, так называемые динамические сети [5]. Текущее состояние инфосюжета может быть представлено в виде графа $\langle M, L \rangle$, где M – это множество информационных сообщений, входящих в сюжет, а L – множество ребер – связей подобия, цитирования, ссылок и т.д. Свойство живучести напрямую связано с такими свойствами графов, как связность, кластерность, минимальный путь и т.п. В инфосюжетах важно выделять основные информационные сообщения, для чего могут использоваться такие методы ранжирования, как PageRank, HITS, Salsa.

На практике задача нахождения и объединения в инфосюжеты отдельных сообщений и их ранжирования предполагает реализацию такой технологической цепочки: построение семантической сети из документов, кластеризацию, "взвешивание" инфосюжетов и их визуализацию [6].

При выделении сюжетных цепочек для определения текстуальной близости сообщений, как правило, используются алгоритмы выявления подобных

документов, ставшие уже традиционными в поисковых системах. Для предъявления пользователям инфосюжеты должны быть ранжированы. Основные факторы, влияющие на ранжирование по важности - оперативность и размер сюжетной цепочки. Под оперативностью понимается некоторая функция от времени публикации всех сообщений в инфосюжете, а размер отражает общий интерес к конкретной теме. Во всех этих подходах центральная задача состоит в отождествлении сообщений, относящихся к одному сюжету и выявление "непересекающихся" сюжетов. Для результирующего отображения инфосюжета используются отобранные по содержательной близости сообщения из различных источников, отсортированные в хронологическом порядке.

Например, в системе Яндекс.Новости (<http://news.yandex.ru>) для построения инфосюжета строится матрица попарной близости документов, которая обрабатывается алгоритмом кластеризации. Для того, чтобы увеличить связность крупных сюжетов, в Яндекс.Новости дополнительно используется кластеризация второго уровня, обеспечивающая сбор атомарных кластеров в более крупные. Ранжирование новостей в инфосюжете основано на количестве и ранге отдельных сообщений.

В системе InfoStream (<http://infostream.ua>) тематическая близость документов определяется на основе нормированных последовательностей наиболее весомых терминов, входящих в каждый документ [7]. Последовательности подобных документов образуют цепочки, которые «взвешиваются» по длине и оперативности. Для отображения сюжетной цепочки, заголовки документов также ранжируются по близости к сюжету, затем из всех заголовков отбираются наиболее "весомые".

Жизненный цикл информационных сюжетов может описываться, например, моделью диффузии информации [8]. В рамках данной модели, которая относится к распространению новостей в информационном пространстве, применяется концепция клеточных автоматов с окрестностью Мура [9] и вероятностными правилами распространения новостей по заданной тематике. Предполагается, что клетка может быть в одном из трех состояний: «свежая новость» (клетка окрашивается в черный цвет); новость, устаревшая, но сохраненная в виде сведений (серая клетка); клетка не имеет информации, переданной новостным сообщением (клетка белая, информация не дошла или уже забыта). Правила развития инфосюжета следующие:

- изначально все поле состоит из белых клеток за исключением одной – черной, которая первой «приняла» новость;
- белая клетка может перекрашиваться только в черный цвет или оставаться белой (она может получать новость или оставаться «в неведении»);
- белая клетка перекрашивается, если выполняется условие: $Cpm > 1$, где p – псевдослучайная величина ($0 < p < 1$), m – количество черных клеток в окрестности, C – константа ($C = 1,5$ при $m = 1$; $C = 1$ при $m \neq 1$);
- если клетка черная, а вокруг нее исключительно черные и серые, то она перекрашивается в серый цвет (новость устаревает, но сохраняется как сведения);
- если клетка серая, а вокруг нее исключительно серые и черные, то она перекрашивается в белый цвет (происходит забывание сведений при их общеизвестности).

Описанная система клеточных автоматов вполне реалистично отражает процесс развития инфосюжета. На поле размером 40 x 40 состояния системы клеточных автоматов полностью стабилизируются за ограниченное количество

ходов, т.е. на практике процесс – сходящийся, это означает, что практически любой инфосюжет имеет конечное время актуальности.

При анализе жизненного цикла инфосюжетов можно использовать еще два больших класса моделей: булевы и марковские.

В булевой модели можно предположить, что инфосюжет состоит из n элементов (сообщений), при этом i -му элементу соответствует булева переменная x_i , которая может принимать значения $\{0,1\}$, то есть $x_i = 1$, если элемент i активен, и 0 – в противном случае. Состояние инфосюжета определяется структурной функцией его работоспособности (булевой функцией) переменных x_1, x_2, \dots, x_n , которая задается в следующем виде:

$$S(x_1, x_2, \dots, x_n) = \begin{cases} 1, & \text{если инфосюжет активен;} \\ 0, & \text{иначе.} \end{cases}$$

Если активность элемента инфосюжета (сообщения) рассматривать как функцию времени, то состояние i -го сообщения можно рассматривать как случайный процесс $x_i(t)$, принимающий в произвольные моменты времени $t \geq 0$ значения 0 и 1. Для системы определяется вероятность ее безотказной работы по известным правилам [10], [11].

Среди недостатков булевых моделей можно назвать предположение только о двух состояниях компонентов – активности и неактивности. Кроме того, булевы модели не учитывают то, что весьма существенную роль может играть последовательность, в которой отказывают отдельные компоненты. Кроме того, в общем случае характер отказов отдельных компонентов зависит от состояния других компонентов. Это находится в противоречии с предполагаемой независимостью элементов в булевой модели.

Инфосюжет можно описать также марковской моделью. Пусть система имеет m возможных состояний. Обозначим множество состояний через $M = \{z_1, z_2, \dots, z_m\}$. Для любого фиксированного момента времени $t \geq 0$ состояние системы $z(t)$ интерпретируется как случайная величина. Заданы множество всех состояний M , вектор распределения начальных вероятностей $p(0)$ и функция переходных вероятностей. Определяется вероятность активности, «жизни» системы в заданный момент времени t (готовность системы) [10]. Применимость марковских моделей также имеет свои границы. Интенсивности переходов между отдельными состояниями системы могут быть нестационарными, принимаемые при расчете допущения относительно распределения интенсивности отказов могут значительно снизить точность полученных результатов; число состояний системы может быть так велико, что расчет становится практически невозможным.

Оценка живучести инфосюжетов должна проводиться на всех этапах их жизненного цикла. Существует несколько подходов, к проведению оценки живучести, имеющих наиболее общий характер. Живучесть системы можно оценить относительно некоторого стандартного внешнего воздействия либо относительно множества внешних воздействий [12]. В этом случае решается задача нахождения множества характеристических векторов состояний инфосюжета, в которых реализуется конфигурация, обеспечивающая выполнение цели функционирования. *Мощность этого множества может служить мерой живучести системы.*

При анализе живучести инфосюжетов рассматривается проблема информирования по их различным аспектам, независимо от наличия или

отсутствия неблагоприятных факторов. В связи с этим, в качестве количественного критерия оценки живучести целесообразно использовать отношение количества функций, выполняемых системой при наличии определенных неблагоприятных воздействий либо множества таких воздействий, к общему количеству функций системы, с учетом критичности выполняемых и не выполняемых функций. Критичность каждой конкретной функции определяется индивидуально для каждого конкретного инфосюжета исходя из его специфики. *Количественный показатель живучести конкретного инфосюжета* в заданных условиях можно вычислять по формуле: $S = \sum_{i \in \Delta} \alpha_i / \sum_{j \in \Theta} \alpha_j$, где Θ – множество

всех функций информирования, Δ – множество функций инфосюжета, выполняемых в заданных условиях ($\Delta \subseteq \Theta$), α_n – критичность n -й функции. Таким образом, количественная оценка живучести инфосюжета будет изменяться в интервале [0, 1], живучесть тем выше, чем больше ее количественная оценка.

Понятие *живучести* инфосюжета подразумевает его способность своевременно выполнять свои функции (в данном случае – информирования) в условиях действия дестабилизирующих факторов. Такими факторами могут быть устранение отдельных сообщений из информационного пространства, потеря их актуальности, доступности. Необходимо отметить, что привлечение внимания аудитории к другой теме, порождение другого информационного сюжета также может снизить актуальность текущего инфосюжета.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Shelton, C., Koopman P., Nace W. A framework for scalable analysis and design of system-wide graceful degradation in distributed embedded systems // Eighth IEEE International Workshop on Object-oriented Real-time Dependable Systems (WORDS 2003), Guadelajara, Mexico, Jan. 2003. -8 p.
2. Додонов А.Г., Кузнецова М.Г., Горбачик Е.С. «Введение в теорию живучести вычислительных систем». Киев: Наук. думка, 1990. –184 с.
3. Горбулін В.П., Додонов О.Г., Ландэ Д.В. Інформаційні операції та безпека суспільства: загрози, протидія, моделювання: монографія / – К.: Інтертехнологія, 2009. – 164 с.
4. Modeling the Revolution in Military Affairs, Autumn // Winter 1998-99 / JFQ.
5. Survivable Network Systems: An Emerging Discipline CMU/SEI-97-TR-013 ESC-TR-97-013 Software Engineering Institute Carnegie Mellon University Pittsburgh, May 1999.
6. Ландэ Д.В. Поиск знаний в Internet. Профессиональная работа – М.: "Диалектика", 2005. – 272 с.
7. Григорьев А.Н., Ландэ Д.В. Адаптивный интерфейс уточнения запросов к системе контент-мониторинга InfoStream // Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции "Диалог"2005" (Звенигород, 1-6 июня, 2005 г.) / – М.: Наука, 2005. – С. 109-111.
8. Ландэ Д.В. Модель диффузии информации // Информационные технологии и безопасность. Менеджмент информационной безопасности. Сборник научных трудов Института проблем регистрации информации. – Вып. 10. – 2007. – С. 51-67.
9. Нейман Дж. Теория самовоспроизводящихся автоматов – М.: Мир, 1971. – 382 с.
10. Райншке К. Модели надежности и чувствительности систем. – М.: Мир, 1979. – 454 с.
11. Райншке К., Ушаков И. А. Оценка надежности систем с использованием графов. – М.: Радио и связь, 1988. – 208 с.
12. Додонов А.Г., Флейтман Д.В. К вопросу безопасности информационных систем // Збірник наукових праць «Інформаційні технології та безпека» - 2004. - Вип.6. Київ. – С. 26-29.